

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Takanori OKURA
Kakuji TORIGOE
Masahi KURIMOTO

(ii) TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE OF INDUCING THE PRODUCTION OF INTERFERON- γ

(iii) NUMBER OF SEQUENCES: 35

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: BROWDY AND NEIMARK
(B) STREET: 419 Seventh Street, N.W., Suite 300
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20004

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Patent In Release #1.0, Version #1.30

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: JP 185,305/96
(B) FILING DATE: 27-JUN-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: BROWDY, Roger L.
(B) REGISTRATION NUMBER: 25,618
(C) REFERENCE/DOCKET NUMBER: OKURA-1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 202-628-5197
(B) TELEFAX: 202-737-3528

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 157 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn
1 5 10 15
Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp
20 25 30
Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg Thr Ile Phe Ile
35 40 45
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
50 55 60
Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys Glu Asn Lys Ile
65 70 75 80
Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys
85 90 95
Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys
100 105 110
Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu
115 120 125

Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu
 130 135 140
 Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp
 145 150 155

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1120 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: No
- (iv) ANTI-SENSE: No
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: liver
- (ix) FEATURE:
 - (A) NAME/KEY: 5'UTR
 - (B) LOCATION: 1..177
 - (C) IDENTIFICATION METHODS: E
 - (A) NAME/KEY: leader peptide
 - (B) LOCATION: 178..285
 - (C) IDENTIFICATION METHODS: S
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 286..756
 - (C) IDENTIFICATION METHODS: S
 - (A) NAME/KEY: 3'UTR
 - (B) LOCATION: 757..1120
 - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCCTGGACAG	TCAGCAAGGA	ATTGTCTCCC	AGTGCATTTC	GCCCTCCTGG	CTGCCAACTC	60
TGGCTGCTAA	AGCGGCTGCC	ACCTGCTGCA	GTCTACACAG	CTTCGGGAAG	AGGAAAGGAA	120
CCTCAGACCT	TCCAGATCGC	TTCCTCTCGC	AACAAACTAT	TTGTGCGAGG	AATAAAG	177
ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG						225
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met						
-35	-30	-25				
AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAT GAA AAC						273
Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn						
-20	-15	-10	-5			
CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA						321
Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile						
1	5	10				
AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT						369
Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro						
15	20	25				
CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA GAT AAT GCA CCC CGG						417
Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg						
30	35	40				
ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG						465
Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met						
45	50	55	60			
GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT						513
Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys						
65	70	75				
GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC						561
Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile						
80	85	90				
AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA AGT GTC CCA GGA						609

Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly
 95 100 105

CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT	657
His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe	
110 115 120	
CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA	705
Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys	
125 130 135 140	
GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA	753
Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu	
145 150 155	
GAC TAGCTATTAA AATTTCATGC CGGGCGCAGT GGCTCACGCC TGTAATCCCA	806
Asp	
GCCCTTTGGG AGGCTGAGGC GGGCAGATCA CCAGAAGTCA GGTGTTCAAG ACCAGCCTGA	866
CCAACATGGT GAAAATCAT CTCTACTAAA AATACTAAAA ATTAGCTGAG TGTAGTGACG	926
CATGCCCTCA ATCCCCAGCTA CTCAGAGGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG	986
GTAGAGGGTG TGGTGAGCG AGATTGCACC ATTGGCCTCT AGCCTGGCA ACAACAGCAA	1046
AACCTCCATCT CAAAAAATAA AATAAATAA TAAACAAATA AAAAATTCAAT AATGTGAAAAA	1106
AAAAAAAAAA AAAA	1120

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..135
- (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA	47
Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser	
-5 1 5 10	
GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT	95
Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn	
15 20 25	
CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G	135
Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp	
30 35 40	

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..134
(C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AT AAT GCA CCC CGG ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC	47
Asp Asn Ala Pro Arg Thr Ile Phe Ile Ser Met Tyr Lys Asp Ser	
40 45 50 55	
CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT	95
Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile	
60 65 70	
TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT TCC TTT AAG	134
Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys	
80 85	

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..87
(C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GAATAAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG	50
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val	
-35 -30 -25	
GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G	87
Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..87
(C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CT GAA GAT GAT G
 Ala Glu Asp Asp Glu
 -10

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2167 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon + 3'UTR
- (B) LOCATION: 1..2167
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA	48
Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile	
85 90 95 100	
TTC TTT CAG AGA AGT GTC CCA GAA CAT GAT AAT AAG ATG CAA TTT GAA	96
Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu	
105 110 115	
TCT TCA TCA TAC GAA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC	144
Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp	
120 125 130	
CTT TTT AAA CTC ATT TTG AAA GAA GAG GAT GAA TTG GGG GAT AGA TCT	192
Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser	
135 140 145	
ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTCA TGCCGGGCGC	246
Ile Met Phe Thr Val Gln Asn Glu Asp	
150 155	
AGTGGCTCAC GCCTGTAATC CCAGTCTTT GGGAGGGCTGA GGCAGGGCAGA TCACCAAGAGG	306
TCAGGTGTTA AAGACACAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA	366
AAAATTAGCT GAGTGTAGTG ACCCATGCC TCAATCCAG CTACTCAAGA GGCTGAGGCA	426
GGAGAAATCA TTGCACTCC GAGGTGGAGG TTGTGTGAG CCGAGATTGC ACCATTGCGC	486
TCTAGCCTGG GCAACAAACAG CAAAACTCCA TCTCAAAAAA TAAAATAAAT AAATAAACAA	546
ATAAAAAAATT CATAATGTGA ACTGTCTGAA TTTTATGTT TAGAAAGATT ATGAGATTAT	606
TAGTCTATAA TTGTAATGGT GAAATAAAAT ATATAACAGT CTTGAAAAAAC ATCATTAAGA	666
AATGAATGAA CTTTCACAAA AGCAAAACAAA CAGACTTCC CTTATTTAAG TGAATAAAAT	726
AAAATAAAATT AAAATAATGT TTAAAAAAATT CATAGTTGA AAACATTCTA CATTGTTAAT	786
TGGCATATTA ATTATACTTA ATATAATTAT TTTTAAATCT TTTGGTTAT TAGTCCTAAT	846
GACAAAAGAT ATTGATATTG GAACTTTCTA ATTTTAAGA ATATCGTTAA ACCATCAATA	906
TTTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAAT TCCCTCCAAA GTCAGTATAT	966
TTTTAAATC CAGTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC	1026
CAACTGACAT TCATCCTAGG AATGCAAGA TGGTTAATA TCCTAAAATC AATTAACATA	1086
ACATACTATA TTAATAAAAGT ATCAAAACAG TATTCTCATC TTTTTTCTT TTTTCACAAT	1146
TCCTGGTTA CACTATCATC TCAATAGATG CAGAAAAAGC ATTTGACAAA ATCCAATTCA	1206
TAATAAAAAT TCTCAAACCTT GAAAGAGAAC ATCATAAAGG CATCTATGAA AACCTACAG	1266
CTAATATCAT ACTTAAACGAT GAAAAACTGA ATTATTTAC CCTARGATCA AGAATAATGC	1326
AAGCATGTCA GCTCTGCAA CTTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA	1386
CCATACAATA AATAAAAATA AAAGGCACCC AGATAGAAAA GGAAGTCTTT ATTGCGAGAC	1446
AACATGGTTC TTTATGCAGA AAACCGTCAG GAATACACAC ACATGTAGA ACTAATAAGT	1506
TCAGCAAGGT TGCAGGTTGC AATATCAATA TGCAAAAAA CATTGAAGGC TGGGCTCAGT	1566
GGAGATGGCA TGTACCTTTC GTCCCACTA CTTGGGAGGC TGAGGTAGGA GGATCACTTG	1626
AGGTGAGGAG TTTGAGGCTA TAGTGCATG TGATCTTGC TGTGAATAGC CACTGCACTC	1686
GAGCCTAGGC AACAAAGTGA GACCCCCTCT CCAAAAAAAA AAATGGTATA TTGGTATTTC	1746
TGTATATGAA CAATGAATGA TCTGAAACAA AGAAAATTCC ATTCAAGATG GTATTAAAAA	1806
AATAAAATAC AAATAAAATT AGCAAAATAA TTATAAAACT TGTACATCGA AAATTCAAA	1866

GCACCTCTGAG GGAAATTAAA GATGATCTAA ATAATTGGAG AGACACTCTA TGATCACTGA 1926
 TTGAAAATT CATTCAATAT TGTTAAGATA ACAATTGTCC CCAAATTGAT GCATGCATTC 1985
 AATTTAGTCT TCATCAAAAT TCCAGCAGGG TTTTGCAGA AATTGACAAG CTGTACCCAA 2045
 AATGTATATG GAAATGAAAAA GACCCAGAAG AGCAAATAAT TTTTAAAAAA CAAAGTTGGA 2105
 AAAACTTTAC TTCCTAATT TAAAACTTAC TATAAACCTA AAGTTATCAA GACCATTAG 2165
 T 2167

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1334 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
 - (A) NAME/KEY: intron
 - (B) LOCATION: 1..1334
 - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTATTTTTT	TAATTGCAA	ACATAGAAAT	GAATAGCTAC	TTCTTCCAT	TCTGTTTAC	60
TGTTTACATT	GTTCCGTGT	AATGCCAATC	CTCAGATGAA	AAGTCACAGG	AGTGACAATA	120
ATTTCACTTA	CAGGAAACTT	TATAAGGCAT	CCACGTTTTT	TAGTTGGGTT	AAAAAATTGG	180
ATACAATAAG	AATTTGCTAG	GGGTCTATGCC	TCTCTGAGCC	TGCCTTGAA	TCACCAATCC	240
CTTTATTGTG	ATTGCATTAA	CTGTTAAAAA	CCTCTATAGT	TGGATGCTTA	ATCCCTGCTT	300
GTACAGCTG	AAAATGCTAA	TAGTTTACCA	GGTGTGGTGG	CATCTATCTG	TAATCCTAGC	360
TACTTGGGAG	GCTCAAGCAG	GAGGATTGCT	TGAGGCCAGG	ACTTTGAGGC	TGAGTACAC	420
TGTGATGTTA	CCTGTGAATA	GGCACTGCAC	TCCAGGCTGG	GTGATATACA	GACCTTGTCT	480
CTAAAATTAA	AAA	AAAAAAACC	TTAGGAAAGG	AAATTGATCA	AGTCTACTGT	540
GCCTTCCAAA	ACATGPAATTC	CAAATATCAA	AGTTAGGCTG	AGTTGAAGCA	GTGAATGTGC	600
ATTCTTTAAA	AATACTGAAT	ACTTACCTTA	ACATATATT	TAAATATT	ATTAGCATT	660
TAAAAGTTAA	AAACAACTT	TTAGAATTCA	TATCTTAAA	TAATCTAAA	AAGTTGCAGC	720
GTGTGTGTTG	TAATACACAT	TAAACTGTTG	GGTTGTTGT	TTGTTGAGA	TGCAGTTTCA	780
CTCTGTCAAC	CAGGCTGAAG	TGCAAGTGCAG	TGCACTGGTG	TGATCTCGGC	TCACTACAAC	840
CTCCACCTCC	CACGTTCAAG	CGATTCTCAT	GCCTCAGTCT	CCCGAGTAGG	TGGGATTACA	900
GGCATGCACC	ACTTACACCC	GGCTAATT	TGTATTTTA	GTAGAGCTGG	GGTTTCACCA	960
TGTTGGCCAG	GCTGGTCTCA	AAACCCCTAAC	CTCAAGTGT	CTGCCTGCCT	CAGCCTCCCC	1020
AACAAACAAA	CAACCCACCA	GGTTAATATG	TGTTACAACA	CACATGCTGC	AACTTTTATG	1080
AGTATTTTAA	TGATATAGAT	TATAAAAGT	TGTTTTAAC	TTTTAAATG	TGGGATTACA	1140
GGCATGAGCC	ACTGTGCCAG	GCCTGAACGTG	TGTTTTAAA	AATGTCTGAC	CAGCTGTACA	1200
TAGTCTCCTG	CAGACTGGCC	AAGTCTCAA	GTGGGAACAG	GTGTATTAAG	GACTATCCTT	1260
TGTTAAATT	TCCGAAATG	TGCTGTGCA	AGAATTCTTC	TAACTAGAGT	TCTCATTAT	1320
TATATTTATT	TCAG					1334

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4773 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 1..4773
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTAGACTGA GCCTTACTTT GTTTTCAATC ATGTTAATAT AATCAATATA ATTAGAAATA
 TAACATTATT TCTAATGTTA ATATAAGTAA TGTAATTAGA AAACCTCAAAT ATCCTCAGAC 120
 CAACCTTTG TCTAGAACAG AAATAACAAG AAGCAGAGA CCATTAAAGT GAATACCTAC 180
 TAAAAATTAT CAAACTCTTT AACTTATGTC ATAATGATGG TTTTCTGAG CCTGTCAG 240
 GGGAAAGAGA GATACACAC TTGTTTATG AACTGATCT CCTGASCAT CAGTCCTTAT 300
 ACAAATAATA ATGTAGAATA CATACTGTAG TTATACATT AAGAATAACA TGTGACTTTC 360
 CAGAATGAGT TCTGCTATGA AGAATGAAGC TAATTATCCT TCTATATTTC TACACCTTG 420
 TAAATTATGA TAATATTTC ATCCCTAGTT GTTTGTCG TGATCCTTAG CCTAAGTCTT 480
 AGACACAAAG TTCACTTTC AGTGTATGTC TGTATTTC AATTTTAATTA TAATTGAAATA 540
 AAAGTTATGA GATCAGCTGT AAAAGTAATG CTATAATTAT CTTCAAGGCCA GGTATAALAGT 600
 ATTTCTGCGC TCTACTTTTT CTCTATTATT CTCCATTATT ATTCTCTATT ATTTTCTCT 660
 ATTTCTCTCA TTATTGTTAG ATAAACACAA ATTACCTATA GCTACAGACT GAGGCACTAA 720
 GAGTAGGCAAG GGATGTTAC AAATTGGCAA TGCTTGAGAG GAGAATTGCA TGTGATGAAG 780
 ACTCTTTTG AGTGGAGATT TGCCATATAA TATCGCTT CATGCCACAG CAGTCCTCAC 840
 TGAAAGAGAG TTAGGATATG ACCTTAATGAG AGGTACCAAG GGGCAACTTG GTAGGGAGAA 900
 AAAAGCCACT CTAACATATA ATCCAAATGAA GAACAGTGCA TATGCAACAG ATACAGCCCC 960
 CAGACAAATC CCTCAGCTAT CTCCCTCAA CCAAGTGCC ACCCTTCAG GTGAAATTTC 1020
 GGAGTCCTCA TTCTAGACCT GACAGCGAGC TTAGTTATCA AAATAGCATA AGAGGCTCTGG 1080
 GATGGAAAGGG TAGGGTGGAA AGGGTAAAGC ATGCTGTTAC TGAAACACAT AATTAGAAGG 1140
 GAAGGAGATG GCCAGCTCA AGCTATGTCG GATAGAGGAA AACTCAGCTG CAGAGCGAGA 1200
 TTCAGAAACT GGGATAAGTC CGAACCTACA GGTGATCT TGTGAGGGAA GACTGTTGAA 1260
 AATGTTAABA AGATGGAAAT AATGTTGGG ACTIAGTAGG AACTGGGAA ATCCATTTT 1320
 GGGGAACTG GAAGTTTATT CAATTTCAT GGGCTTTTA AATAAAAAGA ATGTGCGCTGG 1380
 GGTGCTGCGC TCACACCTGT AATCCCAAGCA CTTTGGGAGG CGAGGGGGGG CGGATCACCT 1440
 GAACTCAAGCA GTCAAGACG AGGCTGACCA ACATGGAGGAA ACCCCATCTC TACTAAAAAT 1500
 ATAAATTAG CTGGCGTGG TGGCATATGC CTGTAATCCC AGTTACTCGG GAGGCTGAGG 1560
 CAGGAGAAATC TTTGAAACCC GGGAGGCAGA GGTGCGATG AGGCTAGATC GTGCGCTGG 1620
 ACTCCAGCTG GGGCAACACG AGCAAAACTC GGTCTCAAAA AAAAAAAMIA AAAAGTAA 1680
 TTAACCAAGG GCATTAGCTT AATAATTAA TACTGTTTTT AAGTAGGGGG GGGGGTGGGT 1740
 GGAAGAAATC TGTTAAATG AGGGAAATCTG ACATTTAACG TTCACTCAGCA TCATGAAA 1800
 TCTGCTGTCG GAGGAAACTC AATAATATT AGTTGGGAGG GGGAGAGAG TGAGGGTGG 1860
 ACTAGGAGCA GTTTTAGCCG TTGTTCTTAA TCTCTTTTCG TGCTCAATAT AAGGATCTTA 1920
 GCACTGTTA TAAAGTGCGC CTAGGTTCTA GATAATATAA TACAAACAGGC CAGTCACAGT 1980
 GGTCTCACTTC TATAATCCCA GCACCTTGGG AGGGCAGGGC GATGTCCTTA CTTGAGATCA 2040
 GGAAGTTCAAG ACCAGCTGG CCAGCATGG CATAATTGTT CTCTACTAAA AAAAAATACAA 2100
 AAATTAACTCA GGCATGTTGG CATCACITG TAATCCAGC TATTCGTGAG CCTGAGGCG 2160
 AAGAATGCT TGAAACCAAG AGGTGTAGG TGAGTGTAGCA CGACTGCACT 2220
 CCAGCTGGG CGACAGAAATG AGACCTTGTG TCAAAAMIA AAAAAAGATAC AACAGGCTAC 2280
 CCTTATGTC TCACCTTCA CTGTTGATTA CTAGTTATAA AGTCTCTATAA AGTTCTTTGG 2340
 TCAAGAACT TGACAACACT AAGAGGGATT TGTTTCTGAG CTTTACTGTC AGAGCTCTTT 2400
 TCAATATATAT ACATATACAT GTATATATGT ATCTTATACG AGGTTGCGC AGGTTCTCT 2460
 CAGACTTTC AGTGCACCTG GGAGATGTTA GGTCAATATC AACTTTCCCT GTTCTGAGAT 2520
 TCAACCCCTT CTGATGTAAA AAAAAAAAAG AAAAAAGAGAG AGATCCTTTT CCTCTTGGAG 2580
 CACTTAAATT TCAAGGTG GGGCTTGGCA AGTTTGCGT TCTCTGAGAGT CTTGAGGTT 2640
 GGTCTCAAT ATCTTGTGTT CGATTTGTTG CTATTATGGG TGAGGAGTGT CATTGAGGTT 2700
 TCAATGTTAT CTGAAATTGTA TGTATGTTAG TCACTTGTGTT AGACCTGTT TCTGAGCTTCC 2760
 TCAACTTGTGTT TTCTTTTTT CTTTTTTTT GAAAGGGAGT CTGGCTCTGT CGGAGGAGGT 2820
 AGAGTGTAGT GGCAGGATCT CAGCTCACTG CGAGTCTGTCG CTGGCTCTGT CGGAGGAGGT 2880
 TCTCTGTTCA CGCTTCCCAAG CAGCTGCGAC TCACTTGTGTT AGTCTGTTGTA TGTTACTAGA 2940
 GGTCTTGTGTT TTCTGCTGTT GGATTATTC TCACTTGTGTT CGAGGAGTGT CTCAGCTCTCC 3000
 CAATGTTATG TGTATGTTAG AAGGTCTTGT AGTCTGTTAG CGAGGAGTGT CTCAGCTCTCC 3060
 AGTGGTGTC AATAAGGGAA CTAGTTTATA CAAATGTTG CGATTACAGG CGTGAACCAT 3120
 ATAAAGAAAGA AGGTGCAAGA GTTATGACTA TGTCTGTTG TGTTACTAGA 3180
 ATAAACCGCT GACCTGTCG ATCCAGAGTC TGTCTATGAG TGTCTGTTGTA AGATGTTGAG 3240
 CTATTTAAAT TTTGGAGGGT TCTCTATGAG TGTCTGTTGTA AGGTTGAGGAG 3300
 TGAGGCCAAA ATTAAGTCA AAAATATAC TGTCTGTTG TGTTACTAGA 3360
 GGTGGTGGCG ACATGTTCTG ATCCATCTG CAAATAGAG ATGCTGCACT AACAGGAAAAA 3420
 TGTCTGTTGAG TGTCTGTTGTA CGAGGAGTGT CTCAGCTCTCC 3480
 TGTTTGGGAT CATGGGAATC TGTCTGTTGTA CGAGGAGTGT CTCAGCTCTCC 3540
 TGTTTCCAGG ATTAACATGG AAGAACCTTA CAAATAGAG ATGCTGCACT AACAGGAAAAA 3600
 GGTGGTGGCG ACATGTTCTG ATCCATCTG CAAATAGAG ATGCTGCACT AACAGGAAAAA 3660

GTGCAGGCAG CACTACCAGT TGGATAACCT GCAAGATTAT AGTTTCAAGT AATCTAACCA 3720
 TTTCTCACAA GGCCCTTATC TGTGACTGAA ACATACAAGA ATCTGCATTG GGCCTTCTAA 3780
 GGCAAGGGCCTC AGCCAAGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAAGTG 3840
 GAGTGTCTGG CAGGAAAGAC AGAGTCAAGG ACTGCCAACT GAGGCAATAC AGCAAGGCTTA 3900
 CACAGGAACG CAGGGCTAG CCTTACAAACA ATTATTGGGT CTATTCACTG TAAGTTTAA 3960
 TTTCAGGCTC CACTGAAAGA GTAAGCTAAG ATTCCCTGGCA CTTCCTGCT CTCTCACAGT 4020
 TGCTCAGAA ATGAGAACTG GTCAAGGCCAG GCATGGTGGC TTACACCTGG AATCCCAGCA 4080
 CTTTGGGAGG CGGAAGTGGG AGGGTCACTT GAGGCCAGGA GTTCAGGACCA AGCTTAGGCA 4140
 ATAAAAGTGAG ATACCCCTG ACCCCCTTCTC TACAAAAATA AATTTAAAAA ATTAGGCAAA 4200
 TGTGGTGGTG TATACTTACA GTGCCAGCTA CTCAGGAGGC TGAGGCAAGGG GGATTGCTTG 4260
 AGCCCAAGGAA TTCAGGCTG CACTGAGCTA TGATTTCAAC ACTGCACTTC TGCTGGC 4320
 ACAGAGGAG ACCCTGTCTC AAAGCAAAA GAAAAGAAA CTAGAACTAG CCTAAGTTTG 4380
 TGGGAGGAGG TCATCATCGT CTTAGCCGT GAATGGTTAT TATAGAGGAC AGAAATTGAC 4440
 ATTAGCCCAA AAAGCTTGTG GTCTTTGCTG GAACTCTACT TAATCTTGAG CAAATGTGGA 4500
 CACCACTCAA TGGGAGGAGA GAGAAGTAAG CTGTTTGATG TATAGGGAA AACTAGAGGC 4560
 CTGGAACCTGA ATATGCATCC CATGACAGGG AGAATAGGAG ATTGGAGTT AAGAAGGAGA 4620
 GGAGGTCAGT ACTGCTGTTC AGAGATTTT TTTATGTAAC TCTTGAGAAG CAAAACACT 4680
 TTGTTCTGT TTGTTAATAT ACTTCAAAAC AAACCTTCATA TATTCAAATT GTTCATGTCC 4740
 TGAAATAATT AGSTAATGTT TTTTCTCTA TAG 4773

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8835 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 1..8835
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GTAAAGAAATA TCATTCCTCT TTATTTGGAA AGTCAGCCAT GGCAATTAGA GGTAAATAAG 60
 CTAGAAAGCA ATTGAGAGGA ATATAAACCA TCTAGCATCA CTACGATGAG CAGTCAGTAT 120
 CARACATAAGA AATATAAGCA AAGTCAGAGT AGAATTTTTT TCTTTTATCA GATATGGGAG 180
 AGTATCACTT TAGAGGAGAG GTTCTCAAAC TTTTGCTCT CATGTTGCTT TTACACTAAG 240
 CACATCACAT GTTACATCAA GTAACATTT TAATTAAAAA TAATCTATGTA CTTTTTAAC 300
 ASGAAAGAAA AGCATAAAGA GTGACACTT TTATTTTTA CAAGTGTGTTT AACTGGTTTA 360
 ATAGAAGGCA TATAAGATCTG CTGGATTCTC ATCTGCTTG CATTGAGACT ACTGCAATAT 420
 TGCACAGAAAT GCAGCTCTG GTAAACTCTG TTGTACACTG ATGAGAGAAAT GGGTGAAAAA 480
 GACAAATTAC GTTCTAGAAT TATTAGAAAT AGCTTCACT TTAGGAACTC CCTGAGAATT 540
 GTGCTTTAG ACTGTTAAGA TAAATAGCT TCTCTTTAA CGGAATCTCA AGACAGAAATC 600
 AATTAACATTA AAGGAAACCA AAAAATTTCG CCATGGTTAG TCATCTGTG AATCTGCCA 660
 CACCTTTGCA CTGGGCTACA ATTGGATAAT ATAGCATTCC CGAGAGATAAT TTTCTCTCAG 720
 AATTAAGGAA AGGGCTGAAAT AAATATCTCT GTTGAAGTT GAATAACAAA ATTAGGAAAC 780
 CGCTAAATTTC TAGGGCTCTG GAAATTCTGC TTTTGCCCTA TATTGAGCTA CTTTACGTT 840
 TATTAATCTCT TCTTCAGGC CAGGTGCACT AGCTCATGCC TAGAATCTCA GGCAGGCTG 900
 AGCCCAGGAA TTGAGAGACCA GCCAGGGCAA CACAGTCTCT AAAAAAAAT AAAAAATTAC 960
 CTGGGTGTGT TGGTGCATGC CTGTAGAAGT ACTCAGGATG CTGAGGACTG CTTGAGGCCA 1020
 GATAGGCCA ATCTGTGGTG AGTTCAAGCC CTAACAGAG CGAGACTTTC TCAAAAAAAAC 1080
 AAACAAAAAA AAAAAACAAA TTCTTCAAA ATAATTTT ATCTGAAATG TTTTCTTATT 1140
 GCTGTGAGA TTAATTTTACCT TCTTTTACCT GATTCCAAA GGCCTCCATA ATCTAATCCG 1200
 ATTTTACCTT GTGTTCACTG CAAAATAGCA GGACTGTTCC ACTACAATCC AAAAATCACA 1260
 GTTGCTCTGC AGTGGCTCAC TCTCTGAAATC CCAACACTTT GGAAGGCCAA GGCAGGCTG 1320
 TTGCTTCAGC TCAGGAGTTA AAGACCAGCC TGGGCAACAT GGCAAAACCT CTGTCCTCC 1380
 AAAACATACA AAAATTAGCC AGATGTGGTA GTATGTGCCT GTAGTCCCAA CTACTCAAAA 1440
 GGCTAAGGCA AGAGGATCAC TTGAGCCAG GAGGTCAAGG CTACAGTGAG CCATGTTAC 1500
 TGTGTCACTG CACTCCAGCC TGGGTGATAG AGCAAGACCA TGTCTCAAAA AAAAAAAAC 1560

GAAAAAAGAAA GAAAAAAACA TGGTCTATT CAGTTAACCC CCACCACAAAC ATTGTTTGA 1620
 TTATCACATA AATGCTGGTC CATTGCCCTTC TCTATCTATT CAAATCTTTA AGGATTCTTT 1680
 GAGATTCAAC TCAATTCTCC TTTCAAAACT AGGCAATTAA AACTACATCA GTTCCATTCTT 1740
 GATTTCTTG CTTTGAGCTT ATAGACTCAA AAACAAAAAC TTAAAAAATT ATTCTTTAAG 1800
 TTTCTGCTA CTCTCACTTC TTCAACACTC ACATACACCG ATTCTATAATA ABATGGCAGA 1860
 ATGTTAAGG ATAAAATGAT TTATAGAACT GAAAAGTTAG GTTCTGATCT TCTTGCTGTC 1920
 AAGATGCTA CCTACCTGAT CTAGGTAAT TAATTATGTA GCATGCTCC TCAATTCACT 1980
 CCATACTAT TCAACAGGGAT TCGAATTCCA CAGCAAGGGAT AAACATAATC ATAGTTGCTT 2040
 TTCAAGTTCA AGGCATTTTA ACTTTTAATC TAGTAGTATG TTTGTTGTTG TTGTTGTTGT 2100
 TTGAGATGGA GCGCTGCTGT GTCACCGAGG CTGGAGTGCA GTGCGACGAA CTGGGTCTAC 2160
 TGCAACCTCT GCCTCATGGG TTCAACTGAGT TATTCTGCCT CAGTGTCCA AGTAGCTGGG 2220
 ACTACAGGAC ACATGCCACC ATGCTGGCT AATTTTGTA TTTTGTAG AAACAGGGCT 2280
 TCACCATGTT GGCCAGGCTG GTCTGAACT CCTGACCTCA AGTGATCAG CGCGCTCGGC 2340
 CTCCCAAAGT GCTGGGATTA CAGGCATAAG CCACCGTGC CAGCCTAATA GTATGTTTT 2400
 AACTCTTAG TGGCTTAACA ATGCTGGTGT TATAATAAAAT ATGCGATAAA TATTTACTGT 2460
 CTTAGAATTA TGAAGAAGTC GTTACTAGC CGTTGCGCAC ATATCAATC TTCTCTGCTT 2520
 ACAGCTTAA TTAGAGTCTA GAATTGAGG TTGTTAGAGC CTTAAAATTA AGTAGCTTC 2580
 GACTAGCAA CTTCCCTGTC CAAATGAGGG AACAGAGACG TGAAATTCTA CCATTCAGTA 2640
 CCCAGACAA ACTGGAACCTC ATGCTGCTA ATTTCCATCA TAAATGAGG TGATGCTGTT TTAAAGGACA 2700
 GCGCTGGCT AGTTCTCAAA GTATTGCGATA ACTAAATTAA TATGCTGTT TTAAAGGACA 2760
 AATTGTCACT GCTTACTCCT GGGAGGCTT TTCTGAGGTG GTTATAACT CTTAAAGAAA 2820
 AAAAAGTCAG TAGTCTGAGA ATTTTAAAGG AAATAGTCAA ASCATTTTA TCCAATEGGAT 2880
 CTATAATTCTT CATAGATTAG AGTTAAATCA AGAAACACG GATGAGAAAG GAAGAGGAAA 2940
 ATTGAGGAGA GGAGGAATGG GGATGAGAAC ACACACTCTG TAATCAGTC TAGATGTA 3000
 GAGAACTAAC AAGAGGAAAT GTAAAGGAAAT AGGAATGAG AATTCAAAT CAAACATGA 3060
 AATAAAGA AACTACTAGG GAAAATGGA GAAGACATTA GAAAATAT TCTATTTTA 3120
 AATTCTGTT TTCAGGCTC CTCCTGTC TTCCCTCTTC CCCTTTCTAC GTGTTGTC 3180
 GAAAATTTA AGATGAAAAA ATATATATAA TTCTCACACAT CATTATGTA GTTCTTACTG 3300
 TGGCAACAGG GTTATCATA GCAAACTTT ATTCTACAA AAGGAGGAACT CATTGCTTTA 3360
 TGTGTTAAGC TCTTCCAGG TGTGTTAAAT TCAGGGGAAA GAAAGACAAT CATTGCTTTA 3420
 AACTCAGAT GAAAGCTGAA CAGACCTATT TTTAATCATA TAGAGCTAT TTAAAGGCA TGAAGGGAGGT 3480
 GTTAGCTAGG AATGAAAGAG CAGAGAGGTT TTCCTAGAGG AACTATTAA GTGTTGAGTT 3540
 ACGGGATGAA AGATGAGGCA GGGTTTGGAG GAAAATGAA AAGGAGGCA GGGAGAGGG 3600
 AAGTTCTGGC CTGGCAGAGA GAAATGAGGT GCGACAACTG AAGGAGGCTC TGGAAAGCAAG 3660
 AAAACCAAGT AGAGAGAAT TAAATGAGA GATGCCAGGG GTATGAGGG CTTGATTAA 3720
 ATGAGTGTGTT TTGAGGATGG AGAGGAGATA CCAAATTGTC GAGCATTTC TGAGTTAGAA 3780
 CCTACAGTAT TTATCAGATA AAGGAAATAT TAGACAAAGT AAGGAAATAT GACTCCCAAG 3840
 TTTCACTTTG GGCGAGGTA CTAGGACATG TTTTGAAAAG TAATGTTTG GATCTCTTAC 3900
 CATTGAAACT ATGTTATGTT AGCTAAATAA AAATTGTA AGACTTCTCA CTTTGTAGCC 3960
 ACCACCAAGTA ACTACTTCCC TAACTCTCA TTGTTCTTCC CATTCTGTC 4020
 TTGTTAGTCAC TGTCTTTACT TTTCCCTCTC CATTCTGTC 4080
 ATCTGCTGCC TCCACCTTAC CGAACTGTT CTAOGGTTGC 4140
 AATTCAATGA ACAAGTTAA GTTATATATG AAATTAGGAG CTCTACAGTT TGTTTCTGAG 4200
 CAGCCCCCTC TGAAACCCCTT TCTTCTGCA CTTCTGTGAC ACATCTGAGA TTACACAAAC 4260
 TGAACTAATT ATTTTACACT TGAGCTGTT TTGTTGTTCTT CTTTCTTCTAT GAATGAGGTA 4320
 ACCACTCAAC AAATGCCCCA AAGGAAAC TACAGGCTA TCTCTAGTT CTCTCTCTTC 4380
 TGTGTTGAGG ACAACAGAT GCTGAGGAA TCCCGCTGTT TAGTATCTCT TGAATTGATT 4440
 ACCTTAATTAT ATAGCTCAT CAACTTTAA TTGTTAAAAT TACTCTGAGTA GTTGTGTTCTT 4500
 GACGCTGTC CAATCTGTT CAATGAGGTC CATTCTTTG TTCTTGTTGG TGTTGTTGTT 4560
 GTTGAGGAGG TGTTGTTTT GCTGTTGAGG CTGAGGTC GAATGAGGACT TCACTGAAAC 4620
 CAACTGCTGTT TGGTTTAAAG CAACTGACCC TCCGGAGTAG CTGGGAGTAC AGGTATGTG 4680
 CAACTGAGGAGG AGCTAATTGTT GTGTTTCTGAG TAGAGACAGG GTTTCAGCAT GTTGTGTTGAG 4740
 CTGGTCTCAA ACTGCTGAGC TCAAGGAACT CAACTGAGGAGG AGCTGAGGAA AGTGTGTTGAA 4800
 TTATGAGGAGT GAGGAGCTGC AGAGGAGGAA GATGCTATTGT TTATGTTGCT TCTATGAGGAA 4860
 GTTCTTAAAG CAACTGAGGAGT CAACTGAGGAGT TCTGAGGAGT AAGGAGGAGT TTTTCTTTC 4920
 AGGAAAGAAAC AGTTCAGGACT CTTTCACTG CTTACACAGG GCGTTTGTAG TGAGCTTCTT 4980
 CTTTCAAGG TTTCATCAAA GATACTGCA AGTACATTT TATGTTGAAAT GAAATTGAGCA 5040
 AGGTTATAAA AATTATAGTT TATATGAGGAA AAATGGAAT ATGTTAACT CTTGAAATAA 5100
 GTTCTTCTAA ATGAGCATAA TTGTTAAAGT GTGAGGAGT AATGAGTATA AACTTTAAC 5160
 ACTATTGCCA CATGCAAGTG TCTTCTTAC TTGTTGAGGAT TATCTCTTC CAGTGTCTA 5220
 TTATGAAAT ATGAGTTAA AATACATGAA AGTCTTCAAG ACATACAGGT TATTATGTTG 5280
 CTTTCTAAAC ATAATTTAA ATAATTTAT ATATGATAAT GTTATCCAAG TCTAAAGGGA 5340
 TGAGTGTGTTA CTGCTGTC AAAAAAAGGAA AAAAAAAGGAA TCCAAATAAA TATGTTGAAA 5400
 CCAAGTTTAT ATGAAAGAAA ACAATATTAA AAAGGCCAAA GTACCACTT AATAAGGCTGT 5460
 GTGAGGAGGAGG CAGGCTACAA AACACTAATA ATAATGCTGA GAAAGTTGAA AAAAGGAAAGA 5520
 AAGGAAACAT ATGCTTTGGT TGTGTTGAGGT TTATGTTACTC CAAAGAATATC TCCTCTCAAA 5580

CTTTTA³TT TTTCCAAAG AAAAGTTAAC TTTGCTGGG CGCACTGGCT CTTGCTGT 5640
 GTCCGAGCT TTGGGAGGCC AAGGCCGGCA GATCACCTGA GGTCAAGGAGT TTGAGACCAG 5700
 CCTGACCAAA AATGGAGAAA CCAGGCCCCC TCACTACTAA AAGAATACAA AATTAGGCG 5760
 GGCACAGTGG CTTACCCCTG TGTACCCAGC ACTTTGGGAG GCGGAAGCAG GAAGATCACC 5820
 TGAGGTCAAG AGTTGAGAC CAGGCACTGA GAAACCCGTC TCTACTAAA ATACAAAATT 5880
 AGCAGGGCTG GGTGCTCAT GACTGTAATC CCAGCTACTC AGGAGGCTAA GGCAGAGAAT 5940
 CACTGAAAC CAGGCACTGG AGTTGCACT GAGGGAGAT CGTGCTATTG CACTCCAGCC 6000
 TGGGCAACAA GAGCGAAACT CTGTATCCAA AAAACAAAAG AAAAGAAAAG GTAACCTTGA 6060
 ACTATGTAG ATCTTGTAAA ATGCATTCTT TCTGTAAAAT GTGACTACAT TTGCTTTATT 6120
 TATGGTAAAAT ATGTTGAGGC CTCAAACAAAC CCATATTTC TCGGTCTCCC CGCTGCCTAG 6180
 CCTTGTTCAT CATTGCTTCT TCTTGGTGG AGCTCTTCCT CTGGCTTGA AAATGCTCTG 6240
 TTCTCTTCA AGGTAGCACA GTCACTCACT TCTGTGGTAA CCTTCTCCAG CACCATCAA 6300
 CAAGAAAGAT GAATCTCTG TAAATTCAAGC TCTTACGTCA TTCATTACAT TATTTGTAA 6360
 CTCTTATAG ATTCTTCTCT CCACACTAGAC TCTGAGTCAC TGGAGAGTAG GAGCCAACCTC 6420
 TCATTCACTGT GTGGTTGCT CAGCTACTGG CCACATTCCT GATGCATAGT TAATGCTCAA 6480
 AGCTTAACG GTGAATCAGC TAAATATTG TCTCTCTCTA AATCCATCTCA CTCATTGACT 6540
 AATATGTAC TCAAAATAGT AACACCCAGT AATTAAATCC AATTCCCTGCC CATACTGCTT 6600
 GGTACATTTC AGGTGATTA GTTGATAAA TATGTGTGA TTACATAATA TTAAAGTATG 6660
 TAAGAAGAT CATGCTAACT ATAATTCAAC ACTGATAACT AATCAAACAT AAATGCTCTC 6720
 AGCTTAACAA ATGTCCTGCCT TCTAGTAA TGCAGTCATT AACAAACACC TTCTGATGCT 6780
 GATAATAGGG CCTTGTTCAG CAATGAAGGC ATAAAGGTGA ATAAAGAACA TGCCCTCGTG 6840
 GAAGTCACAG CCTAGTCATT ATTGTTCTGA TTTTAATAT TAATGTTGGT TTGGGTTTTG 6900
 GTGAAAGATG TTTAGACTTA TCTTAGTGTAT CTTTCATCC TTTGCTATAT TATTTTCTC 6950
 TAAGAGTCTT CCTTATCCCC TCTTTAAAAA AACTAGGTGA TAATTCTAAA TTGTAATTT 7020
 AATATTATA AATAGCTTAT AAAATTTAAT ATTATAATAA TTTAAATGTT TGATAAATAT 7080
 TTAAATTTA TAATATTTA ATGTTTATT AAATTCATT AAATCCTAGT TTTTATTTA 7140
 TTAAATCTG TTGGCCAGGC ATGGTGCCTG ACACCTATAA TCCCAGAACT TTGAGAGGCC 7200
 AGTCAGGCA AACCATTGCA GCTCAGGAGT TTGAGACCAC CCTGGGCAAC GTGGTGAAC 7260
 CCTGTCTCTA CCAAACATAT GAAACTTAT CTGGTGTGG TGGCACCGCAT CTGGTGTCCC 7320
 AGATGGGAGT CCCAGGTAA GATGGAGAA TCGCTTGAAC CCAGGTGAGA GGGGTGGGGT 7380
 GGATGTTGCA GTGAGCTGAG ATCGTGTGAC TGCACCTCAA CCGGGTGAC AGAGTGAGAC 7440
 TCCATOTCAA AAAAAGAAA TGTATTTAA ATAAGATAAA TTAACACAA TAAGGCCTGG GTACAGTGAC 7560
 AGATGACCAT AAGGAACCAA ACCTAGATAA AACTATCAA TACAAAGTTA GTTGTATAAC 7620
 TCATGCCTGT AATCTCAAGC ACGTTGGAG GCGAAATTA AATTTTTTTT AACTGAGTT 7680
 ACCAACTAAC AACTATTTG GGGTTAGCTT AATTCAAGATT ATTATGCTAG AAAAATTTG 7740
 TTAAATTCTG GCTTACTCTA CCATACTATGC TAGGCTCTAT TTTAAACAAA TTAAATGCA 7800
 AGCAGAGTT TATGAATACT CTCTGCAATA CCATTAAATT TCTACTGTGA GGACTAAATT 7860
 GTATATATGT GCCTTTTAC CAACACATTA AATAATAAG AAGATCTCTG CAATAACTGT 7920
 TCTGTAATT CAAAGTAGTA ATGAGCTTAA ACCATGTCTC ACAAGTACTA CTAATAATAC 7980
 AGCACAAACAG AAAATAGCTA TTCTCTTAA TGACAGAGTC AATTTCAGAG GTTGGTGAAA 8040
 TGTGGTTGT TTCTCTGAG TAACTCTGG AGGAATGCTA AATCCAGATC AAAAATCACC 8280
 ATACATGTGT ATTTTTCTC CCATCAAGT TCACAGATT CTCACACTGA GAACTCCTAT 8100
 TCCATAACAA AATTCGAAAG CGCTGCACAC CGTATTGGAA GAAGGGCAGA AAGGAAAAGC 8160
 AAATGGGAAAG ATTTAAATTT TTTTAAATC CTGTTATCCCT TGATTTTACA GCAAGATTGT 8220
 ATTTATGTAT TACTTCTTTT AAAAATATAG TATAATCGAG ACTCCAGATC AAAAATCACC 8280
 GAGGCTGAGG GAGAAAGAGG CGCAGAAAT GCGAGAGCCC TTCAGCCTTC TCCCACCCCTG 8340
 CCTGTACCT CAGATGCGAG CAGTTTTTA TCATGTTTC ACCTTAGCA TTTTGACAAAT 8400
 GAAAGTCACAA AGCTTCAGGC TCTCACCCAT AGGAACCCAC TGGTTGTAAG AGAAGGATGA 8460
 AGCCAGTCCT TCTTAAGGG CAGGATTAGA TGTGTTTATG GCTACCTCAG GTGAAACTAT 8520
 ATTTATATTG ACAAATATAG TATATTTCTC AAGGAATACT AGAATAATGA TTCACTGAG 8580
 TACTAGGGCA TTATGTTATC TTCTTATAATG TGTGTTAATG AGAAAATGCT TTCTATCTC 8640
 CAATATTCTG ATGATTGCTA AGAGAACACT TAAACATGGG TATTCATAAG CTGAAACTTC 8700
 TGCGATTTAT TGAATGCTAA GATTGTTCTAT CAGTATACTA GGTGATTAAC TGACCACTGA 8760
 ACTTGAAAGT AGTATAAAGT AGTAGTAAAAA GGTACAATCA TTGTCTCTTA ACAGATGGCT 8820
 CTTTGTCTTC ATTAG 8835

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1371 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 1..1371
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTAAGGCTAA	TGCCATAGAA	CAAATACCAAG	GTTCAGATAA	ATCTATTCAA	TTAGAAAAGA	60
TGTTGTGAGG	TGAACATTAA	AGTGACTCTT	TGTGTCACCA	AATTTCACTG	TAATATTAAT	120
GGCTCTTAAA	AAAATAGTGG	AACCTCTAGAA	ATTAACCACA	ACATGTCCAA	GGTCTCAGCA	180
CCTTGTACAA	CCACGTGTCC	TGGCACTTTA	ATCAGCAGTA	GCTCACTCTC	CAGTTGGCAG	240
TAAGTGCACA	TCATGAAAAT	CCCAGTTTC	ATGGGAAAAT	CCCAGTTTC	ATTGGATTTC	300
CATGGGAAAAT	ATCCCAGTAC	AAAAGTGGGT	GCATTCAGGA	AATACAATTTC	CCCAAAGCAA	360
ATTGGCAAAT	TATGTAAGAG	ATTCCTCTAA	TTTAGAGTTTC	CGTGAATTAC	ACCATTTTAT	420
GTAATATGT	TTGACAAGTA	AAAATTGATT	CTTTTTTTT	TTTCTGTTG	CCCAGCTGG	480
AATGCAGTGG	ACAATCTCT	GCTCACTGCA	ACCTCCACCT	CCTGGTTCA	AGCAATTCTC	540
CTGCCTCAGC	CTTCTGAGTA	GCTGGAACTA	CAGGTGCATC	CCGCCATGCC	TGGCTAATTTC	600
TTGGGTATT	TTACTAGAGA	CAGGGTTTG	GCATGTTGTC	CAGGCTGGTC	TTGGACTCCT	660
GATCTCAGAT	GATCCTCCTG	GCTGGGCTC	CCAAAGTGT	GGGATTACAG	GCATGAACCA	720
CCACACATGG	CCTAAAAATT	GATTCTTATG	ATTAATCTCC	TGTGAACAAT	TTGGCTTCAT	780
TTGAAAGTTT	GCCTTCATTT	AAAACCTTC	TTTAAAAGCC	TGAGCAACAA	AGTGAGACCC	840
CATCTCTACA	AAAAACTGCA	AAATATCCTG	TGGACACCTC	CTACCTCTG	TGGAGCTG	900
AGCAGGAGGA	TCACTTGAGC	CTAGGAATT	GAGCTGCGAG	TGAGCTATGA	TCCCACCCCT	960
ACACTCCAGC	CTGCACTGACA	GTAGACCCCTG	ACACACACAC	ACAAAAAA	ACCTTCATAA	1020
AAAATTATTA	GTTGACTTTT	CTTAGGTGAC	TTTCGTTTA	AGCAATAAT	TTAAAGTAA	1080
AATCTCTAA	TTTAAAGAAAT	TTATTTTAG	TTACATATTG	AAATTTTAA	ACCTAGGTT	1140
TAAGTTTTAT	GTCTAAATTA	CCTGAGAACAA	CACTAAGTCT	GATAAGCTTC	ATTATGGG	1200
CCTTTGGAT	GATTATATAA	TATTCTGATG	AAAAGCCAAGA	CAGACCTTA	AACCATAAAA	1260
ATAGGAGTTC	GAGAAAGAGG	AGTAGCAAAA	GTAAAAGCTA	GAATGAGATT	GAATTCTGAG	1320
TCGAATACAA	AAATTTACA	TATTCTGTT	CTCTCTTTT	CCCCCTCTTA	G	1371

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3383 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 1..3383
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GTAAGTAGA	AATGAATTAA	TTTTCTTTG	CAAACAAAGT	ATCTGCTTGA	GACACATCTA	60
TCTCAACATT	GTCAAGCTGAG	AAAAAAGAAA	AATGGTTCTC	ATGCTACCAA	TCTGCTTCA	120
AAGAAATGTG	GACTCAGTAG	CACAGCTTTG	GAATGAAGAT	GATCATTAAGA	GATACAARGA	180
AGAACCTCTA	GCAAAAGATG	CTTCTCTATG	CCTAAAAAAA	TTCTCCAGCT	CTTAGAATCT	240
ACAAAAATAGA	CTTTGCCTGT	TTCAATTGTC	CTAACAGATTAG	CATGAAGCCA	TGGATTCTGT	300
TGTAGGGGG	GCCTTGCATA	GGAAAAAGGG	ATTGAAGCAT	TAGAATTGTC	CAAATCAGT	360
AACACCTCTC	CTCAAGAAATG	CTTGGAAAG	AAGCCTGGAA	GGTTCGGGT	TGGTGGTGGG	420
GTGGGGCAGA	AAATTCTGGA	AGTAGAGGGAG	ATAGGAATGG	GTGGGGCAAG	AAGACCACAT	480
TCAGAAGGCA	AAAGCTGAAA	GAAACCATGG	CATTATGAT	GAATTCAAGGG	TAATTCAAGAA	540
TGGAAGTAGA	GTAGGAGTAG	GAGACTGGT	AGAGGAGCTA	GAGTGTAAA	CAGGGTGTAG	600
AGCAAGACGT	TCTCTCACCC	CAAAGATGTGA	AATTGGACT	TTATCTGGAA	GATAATAGGG	660
TTAATTAAGC	ACAATATGTA	TTAGCTAGGG	TAAAGATTAG	TTTGTGAA	CAAAGACATC	720

CAAAGATAACA GTAGCTGAAT AAGATAGAGA ATTTTTCTCT CAAAGAAAGT CTAAGTAGGC 780
 AGCTCAGAAG TAGTATGGCT GGAAGCAACC TGATGATATT GGGACCCCCA ACCTCTTCA 840
 GTCTTGTACC CATCATCCCC TAGTTGTTGA TCTCACTCAC ATAGTTGAAA ATCATCATAC 900
 TTCTCTGGTT CATATCCCCAG TTATCAAGAA AGGGTCAAGA GAAGTCAGGTC ATTCCTT 960
 CAAAGACTCT AATTGGAAAGT TAAACACATC AATCCCCCTC ATATTCCATT GACTAGAATT 1020
 TAATCACATG GGCACACGCAA GTGCAAGGAA ATCTGGAAAA TATAATCTT ATTCCAGGTA 1080
 GGCATATGAC TCTTTAAAT TCAAGAAATA TATATTTTA AAATATCATT CTGGCTTTGG 1140
 TATAAAGAAT TGATGGTGTG GGGTGAGGAG GCGAAAATTA AGGGTTGAGA GCCTATTATT 1200
 TTAGTTATTA CAAGAAATGA TGGTGTCAATG AATTAAGGTA GACATAGGGG AGTGTGATG 1260
 AGGAGCTGTG AATGGATTT AGAAACACTT GAGAGAATCA ATAGGACATG ATTAGGGTT 1320
 GGATTTGGAA AGGAGAAGAA AGTAGAAAAG ATGATGCTA CATTTCAC TTAGGCAATT 1380
 TGTACCATTC AGTGAATAG GGAACACAGG AGGAAGAGCA GGTGTTGTTG TATACAAAGA 1440
 GGAGGATGGA TGACGCATTG CGTTTGGAT CTGAGATGTC TGTGGAACGT CCTAGTGGAG 1500
 ATGTCCACAA ACTTTCTAC ATGTGTTCT GAGTCAGGA CACAGATTTG GGCTGGAGAT 1560
 AGAGATATTG TAGGTTATA CATAGAAATG GCATTTGAAT CTATAGAGAT AAAAGACAC 1620
 ATCACAGGAA ATGTGTTAAAG TGAGAGAGGA AAAGCCAAGT ACTGTGCTGG GGGGAATACC 1680
 TACATTAAAGGATGCACTA GAAAGAAGCT AATAAACACAGA AGAGAGCAGA CTAACCAAAA 1740
 GGGGAGAAGA AAAACCAAGA GAATTCCACCGA GATTCCTCAGGAGAAGCATTTC AAGATTGAGG 1800
 GGATAGGTGT TGTGTTGAAT TTTGCACTTGAGAATCAA GGGCCAGAAC ACAGCTTTA 1860
 GATTAGCAA CAAGGAGTTT GGTGATCTCA GTGAAAGCAG CTTGATGGT AAATGGAGGC 1920
 AGAGGCAGAT TGCAATGAGT GAAACAGTGA ATGGGAAGTG AAGAAATGAT ACAGATAATT 1980
 CTTGCTAAA GCTTGGCTGT TAAAAGGAGG AGAGAAACAA GACTAGCTGC AAAGTGAGAT 2040
 TGGGTTGATG GAGCAGTTT AAATCTCAA ATAAAGAGCT TTGTTGCTTT TTGATTATGA 2100
 AAATAATGTG TTAATTGTAA CTAATTGAGG CAATGAAAAAA AGATAATAAT ATGAAAGATA 2160
 AAAATATAAA AACCACCCAG AAATAATGAT AGCTACCAATT TIGATACAAAT ATTCTACAC 2220
 TCCTTCTAT GTATATATAC AGACACAGAA ATGCTTATAT TTTTATTAAAGGATTGTA 2280
 CTATACCTAA GCTGCTTTT CTAGTTAGT ATATATATGG ACATCTCTCC ATGGCAACGA 2340
 GTAATTGCAAG TTATATTAAG TTCACTGATATT TTCACAAATAA GGCATATCT TTGCCCCTTT 2400
 TATTAAATCA ATTCTTAATT GGTGAATGTT TGTGTTCCAGT TTGTTGTTGTTG TATTAACAAT 2460
 GTTCCATTAAGCATTGTCAGT ACACCAATGTT TCACACATT GTCTGATTT TTCTTCAGGA 2520
 TAAACCCAG GAGGTAGAAT TGCTGGTTG ATAGAAAGAGA AAGGATGATT GCCAAATTAA 2580
 AGCTTCAGTA GAGGGTACAT GCGGAGCACA AATGGGATCA GGCCTAGATA CCAGAAATGG 2640
 CACTTTCTCA TTTCCCTTG GGACAAAAGG GAGAGAGGCA ATAACGTGTC TGCCAGAGTT 2700
 AAATTTGTAC GTGGAAGTAGC AGGAAATCAT TTGCTGAAAAA TGAAAACAGA GATGATGTTG 2760
 TAGAGGTCTG GAAGAGAGCA AAGAAAATTG GAAATTGCGG CTATCAGCTA TGGAAGAGAG 2820
 TGTGTAATG GAAAACAAAAA GAAGTATTGA CAATTTGTT GCTTGTATG GCACCGATTT 2880
 GAACGCTTGT GCCATTGTTCA ACCAGCAGCA CTCAGCAGCC AAGTTGGAG TTTGTTGCA 2940
 GAAAGACAAA TAAGTTAGGG ATTAAATATC CTGGCTCAAAT GGTAGACAAA ATGAAACTCTG 3000
 AGATCCAGCT GCACAGGGAA GGAAGCTGAAG ACGGGAAGAG GTTAGATAGG AAATACAAGA 3060
 GTCTGGAGAC TGGAAGATGT TGTTGATATT AAAGAACACAT AAGTTGAGG TAAAAGTGT 3120
 AGAAAAACTAG AAGGTTAAGA GACCGGTCAAG AAGTAGGTT ATTGAAAGTT AACACTTCAG 3180
 AGGCAAGAGTA GTTCTGAAATG GTAAACAGAA ATTGAGTGTG CTTTGAGAG TAGGTAAAAA 3240
 AAATAATAGGC AACTTTATTG TAGCTACTTC TGGAACAGAA GATTGTCAATT AATAGTTTTA 3300
 GAAAACAAAAA ATATATAGCA TACTTATTG TCAATTAAACA AAGAAACTAT GTATTTTAA 3360
 ATGAGATTTA ATGTTTATG TAG 3383

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11464 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
 - (A) NAME/KEY: 5'UTR
 - (B) LOCATION: 1..3
 - (C) IDENTIFICATION METHODS: E
 - (A) NAME/KEY: leader peptide
 - (B) LOCATION: 4..82
 - (C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron
 (B) LOCATION: 83..1453
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 1454..1465
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 1466..4848
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 4849..4865
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 4866..4983
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 4984..6317
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 6318..6451
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 6452..11224
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 11225..11443
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: 3'UTR
 (B) LOCATION: 11444..11464
 (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AAG ATG GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA	48
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala	
-35 -30 -25	
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGG CTAATGCCAT	98
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	
AGAACAAATA CCAGGTCAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAAC	158
ATTAAGTGAC TCTTGTCG ACCAAATTTC ACTGTAATAT TAATGGCTCT TAAAAAAATA	218
GTGGACCTCT AGAAATTAAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCACGT	278
GTCCTGGCAC TTTAATCAGT AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA	338
AAATCCCAGT TTTCATGGGA AATCCCAGT TTTCATGGGA TTTCCATGGG AAAAATCCCA	398
GTACAAAACT GGGTCCATTC AGGAAATACA ATTTCCAAA GCAAATTGCG AAATTATGTA	458
AGAGATTCTC TAAATTAGA GTTCGGTGA TTACACATT TTATGTAATT ATGTTGACA	518
AGTAAAAAATT GATTCTTTT TTTCCTTGT GTTGCCTCAGG CTGGAGTGCA GTGGCACAAAT	578
CTCTGCTCAC TGCACACTCC AACTCCTGG TTCAAGCAAT TCTCCTGCCT CAGCCTCTG	638
AGTAGCTGGG AATACAGGTG CATCCTCGCA TGCCCTGGCTA ATTTTTGGGT ATTTTTACTA	698
GAGACAGGGT TTTGGCATGT TGTCCAGGCT GGTCTGGAC TCCTGATCTC AGATGATCT	758
CCTGGCTCGG GTCGCGAAAG TGCTGGATT ACAGGCATGA ACCACACAC ATGGCCTAAA	818
AATTGATTCT TATGATTAAT CTCTGTGAA CAATTGGCT TCATTGAAA GTTGCCTTC	878
ATTGAAACC TTCAATTAAA AGCTGAGCA ACAAAAGTGAG ACCCCATCTC TACAAAAAAC	938
TGCAAAATAT CCTGTGGACA CCTCTCTACCT TCTGTGGAGG CTGAAGCAGG AGGATCACT	998
GAGCCTAGGA ATTGAGGCT GAGTGAGCT ATGATCCAC CCCTACACTC CAGCCTGCAT	1058
GACAGTAGAC CCTGACACAC ACACAAAGAAA AAAAACCTTC ATAAAAAATT ATTAGTTGAC	1118
TTTCCTTAGG TGACTTTCCG TTTAAGCAAT AAATTAAAAA GTAAAATCTC TAATTTTAGA	1178
AAATTATTAT TTAGTTACAT ATTGAATT TTAAACCTA GGTTAAAGTT TTATGTCTAA	1238
ATTACCTGAG AACACACTAA GTCTGATAAG CTTCATTITA TGGGCCTTT GGATGATTAT	1298
ATAATATTCT GATGAAAGCC AAGACAGACCT CTTAAACCAT AAAAATAGGA GTTCGAGAAA	1358
GAGGAGTAGC AAAAGTAAAAA GCTAGAAATGA GATTGAATTC TGAGTCGAAA TACAAAATT	1418
TACATATTCT GTTCTCTCT TTTCCCTCT CTTAG CT GAA GAT GAT G GTAAA	1470
Ala Glu Asp Asp Glu	
-10	
GTAGAAATGA ATTTATTTT CTTGCAAAC TAAGTATCTG CTTGAGACAC ATCTATCTCA	1530
CCATTGTCAG CTGAGAAAA AAAAATGG TTCTCATGCT ACCAATCTGC CTTCAAAGAA	1590
ATGTGGACTC AGTAAACACAG CTTGGAAATG AAGATGATCA TAAGAGATAC AAAGAAGAAC	1650

CTCTA GCAAA AGATGTTCT CTATGCTTA AAAAATTCTC CAGCTCTTAG AATCTACAAA	1710	
ATAGACTTTG CCTGTTTCAT TGGTCCTAAG ATTAGCATGA AGCCTATGGAT TCTGTTGAG	1770	
GGGGAGCGTT GCATAGGAAA AAGGGATTGA AGCATTAGAA TTGTCCAAGA TCAGTAACAC	1830	
CTCTCTCAAG AAATGTTTG GAAAGAAGCC TGGAGGTTG CGGTTGGTGTG GTGGGGTGGG	1890	
GCAGAAAATT CTGGAAGTAG AGGAGATAGG ATGGGTGGG GCAAGAAGAC CACATTAGA	1950	
GGCCAAAAGC TGAAAAGAAC CATGGTATT ATGATGAATT CAGGGTAATT CAAATGGAA	2010	
GTAGAGTAGG AGTAGAGAC TGGTAGAGG AGCTAGAGTG ATAAACAGGG TGAGAGTAA	2070	
GAGGTTCTCT CACCCCAAGA TGTGAAATTG GAGCTTATC TTGGAGATA TAGGGTTAAT	2130	
TAAGCACAAT ATGTTAGG TAGGGTAAAG AGAGAATTTC TGTAAACAGG ACATCCAAAG	2190	
ATACAGTAGC TGAATAAGAT AGAGAATTTC TGTCTAAAG AAAGCTCTAG TAGGCAGCTC	2250	
AGAAGTAGTA TGCTGGAAAG CAACCTGATG ATATTGGGAC CGCCAACCTT CTTCAGTCTT	2310	
GTACCCATCA TCCCCTAGTT GTTGATCTCA CTACATAGT TGAAAATCAT CATACTCCT	2370	
GGGTTCATAT CCCAGTTATC AAGAAAGGGT CAAGAGAAAGT CAGGCTCATT CCTTCAAG	2430	
ACTCTAATTG GAAGTTAAAC ACATCAATCC CGCTCATATT CCATTGAGTA GAATTAACTC	2490	
ACATGCCCAC ACCAAGTGCAGGAAATCTG GAAAATATAA TCTTTATTCC AGGTAGCCAT	2550	
ATGACTCTT AAAATTCAAG AATAATATAT TTTTAAATAA TCATTCTGGC TTTGGTATAA	2610	
AGAATTEATG GTGTGGGGTG AGGAGGCCAA AATTAAGGGT TGAGAGCCTA TTATTTAGT	2670	
TATTACAGA AATGATGGT TCATGAATTA AGGTAGACAT AGGGGAGTG TGATGAGGAG	2730	
CTGTGAATGG ATTTTAGAAA CACTTGAGAG AATCAATAAG ACATGTTTA GGTTTGAGATT	2790	
TGGAAAGGG AGGAAAGTAG AAAAATGAT GCCTACATT TTCACTTACG CAATTGTTAC	2850	
CATTCACTGA AATAGGGAAAC ACAGGGAGAA GAGCAGGTT TGTTGTATAAC AAAGAGGAGG	2910	
ATGGTAGGAGG CATTTCGTTT TGTGATGAG ATGTCCTGAA AACGTCCTAG TGGAGATGTC	2970	
CACAAACTCTC TGTACATGTT GTTCTGAGTT CAGGACACAG ATTGGGCTG GAGATAGAGA	3030	
TATTGAGGTT TTATACATAG AAATGGCATT TGAATCTATA GAGATAAAAA GACACATCAG	3090	
AAGAAATGTT TAAAGTGAGA GAGGAAAGC CAAGTACTGT GCTGGGGGGG ATACCTACAT	3150	
TTAAAGGATG CAGTAGAAAG AAGCTAATAA ACAACAGAGA GCAGACTAAC CAAAGGGGA	3210	
GAAGAAAAAC CAAGAGAATT CCACCGACTC CCAGGAGAGC ATTCAAGAT TGAGGGATA	3270	
GTTGTTCTGT TGAAATTGTC AGCCTTGAGA ATCAAGGCG AGAACACAGC TTTTAGATT	3330	
AGCAACAAAGG AGTTGGTGA TGTCACTGAA AGCAGCTGA TGTTGAAATG GAGGAGG	3390	
CAGATTCGAA TGAGTGAAC AGTGAATGGG AAGTGAAGAA ATGATACAGA TAATTCTGC	3450	
TAAGAACCTTG GTGTTAAAGA GGAGGAGAG AACAAGACTA GCTGCAAAGT GAGTTGGGT	3510	
TGATGGAGCA GTTTAAATC TCAAAATAAA GAGCTTGTG CTTTTTGAT TATGAAAATA	3570	
ATGTGTAAAT TGTAACTAAT TGAGGCAATG AAAAAAGATA ATATATATGAA AGATAAAAAT	3630	
ATAAAACCCA CCCAGAAATA ATGATAGCTA CCATTITGAT ACAATATTC TACACTCCTT	3690	
TCTATGTATA TATACAGACA CAGAAATGCT TATATTTTA TTAAAGGGG TTGTACTATA	3750	
CCTAAGCTEE TTTTTCTAGT TAGTGATATA TATGGACATC TCTCCATGGC AACGAGTAAT	3810	
TGCACTTATA TTAAGTTCAT GATATTTCAC AATAAGGGCA TATCTTGGC CTTTTTATT	3870	
AAATCAATTCT TAATTCGTTA ATGTTTGTGTT CCAAGTTGTT GTTGTATT ACAATGTTCC	3930	
CATAAGCATT CCTGTACACG AATGTTACAA CATTGTCCTG ATTTTTCTT CAGGATAAA	3990	
CACAGGAEST AGAATTGCTG GTTGTATAGA AGAGAAAGGA TGTTGCGAA ATTAAAGTT	4050	
CAGTAGAGGG TACATGCGA GCACAAATGG GATCAGCCCT AGATACCCAGA AATGGCACTT	4110	
TCTCTTCCAT CCTGGGACA AAAGGGAGAG AGCCTAAAC TGTGTGCGCA GAGTTAAATT	4170	
TGTACCTTGA GTAGAGGAA ATCATTGTT GAAAATGAAA ACAGAGATGA TGTTGTAGAG	4230	
GTCCTTAAAGA GAGCAAGAAA AATTGAAAT TGCGGCTATC AGCTATGGAA GAGAGTGTG	4290	
AACTGAAAAAA CAAAAGAAGT ATTGACAATT GGTATGCTG TAATGGCAGC GATTGAAACG	4350	
CTTGTGCGCAT TGTTCACCG CAGCACTCAG CAGCCAAAGTT TGGAGTTTG TAGCAGAAAG	4410	
ACAAATAAGT TAGGGATTAA ATATGCTGC CAAATGGTAG ACACAAATGAA CTGTGAGATC	4470	
CAGCTGCGACA GGGAAAGCGAG GGAAGACCGGG AGAGGTTAG ATAGAAAATA CAAGAGTCAG	4530	
GAGACTGAAA GATGTTGAGA TATTAAAGA CACATAGAGT TGGAATAGA GTGTAAGAAA	4590	
ACTAGAGGGG TAAGAGACCG GTCAAGAAGT AGCTATTTG AAGCTAACAC TTTCAGAGGCA	4650	
CAGTACTGTT GAATGGTAAAGAAGGATTGA GTGTTGCTTT GAGAGTAGAT TAAAGAAATA	4710	
TAGGCAACTT TATTGAGGTT ACTTGTGGAA CAGAAGATG TGCAATTAAAG TTTTAGAAAAA	4770	
CTAAATATA TAGCATACATT ATTGTCATTAACAAAGAA ACTATGTTATT TTTAAATGAG	4830	
ATTTAATGTT TATTGAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT	4880	
Glu Asn Ile Glu Ser Asp Tyr Phe Gly Lys Leu		
-5	1	5
GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC	4928	
Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe		
10	15	20
ATT GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC	4976	
Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp		
25	30	35
TGT AGA G GTATTTTTT TTAATTGGCA AACATAGAAA TGAGTAGCTA CTTCTCCCA	5032	
Cys Arg Asp		
40		
TTCTGTTTGA CTGCTTACAT TGTCCCGTGC TAGTCCCAAT CCTCAGATGA AAAGTCACAG	5092	
GAGTGACAAT AATTCACCT ACAGGAAACT TTATAAGGCA TCCACGTTTT TTAGTTGAGG	5152	

TAAAAAAATTG GATACAATAA GACATTGCTA GGGGTCTGCT CTCTCTGAGC CTGCCCTTGA 5212
 ATCACAATC CTTTATTGT GATTGCTTA ACTGTTTAAA ACCTCTATAG TTGGATGCTT 5272
 AATCCCTGCT TGTTACAGCT GAAAATGCTG ATAGTTTACG AGGTGTGCTG GCTATCTATCT 5332
 GTAATCCTAG CTACTTGCGA GGCTCAAGCA GGAGGATTGCT TTGAGGCGAG GACTTTGAGG 5392
 CTGTAATGCTA CTGTGATGCT ACTGTGAAAT AGCCACTGCA CTCCAGCTG GGTGATATAAC 5452
 AGACCTTGTCTC TCTAAAATTA AAAAAAAAATTA CTTAGGAAAG GAAATTGATC 5512
 AAGTCTATGCTG TGCCTTCGAA AACATGAAATT CCAATATCA AAGTTAGCTG GAGTTGAAGC 5572
 AGTGAATGCTG CATTCTTAA AAATACGAA TACTTACCTT AACATATATT TAAATATTT 5632
 TATTTAGCAT TTAAAAGTTA AAACAAATCT TTTAGAATTC ATATCTTAA AATACTCAA 5692
 AAAGTTGCAG CGTGTGTGTT GTAATACACA TTAAACTGTC GGTTGTTGTTGAG 5752
 ATGCAGTTTC ACTCTGTCAC CCAGGCTGAA GTGCACTGCA GTGCACTGCTG GTGATCTCG 5812
 CTCACATCAA CCTCCACCTC CCAGGTTCAA GCGATTCTCA TGCTCTAGTC TCCCGAGTAG 5872
 GTGGGATTAC AGGCATGAC CACTTACACG CGGCTAATT TTGTGATTTTT AGTAGAGCTG 5932
 GGGTTTCACC ATGTTGCCCA GGTGCTGCTC AAACCCCTAA CCTCAAGTGA TCTGCTGCG 5992
 TCAGCCCTCCC AAACAAACAA ACACCCCCAC AGTTTAATAT GTGTTACAC ACACATGCTG 6052
 CAACTTTAT GAGTATTITA ATGATATAGA TTATAAAAAG TTGTTTTAACTTTAAATG 6112
 CTGGGATTAC AGGCATGAGC CACTGTCGA GGCCTGAACT GTGTTTTAAATGTTGAG 6172
 CCAGCTGTAC ATAGTCTCTC GCAAGCTGCA CAACTCTCAA AGTGGAAACA GGTGTTAA 6232
 GGACTATCCT TTGTTAAAT TTGCTGCAATT GTTCTGCTGCA AAGAATTCTT CTAACCTAGAG 6292
 TTCTCATTTA TTATATTAT TTGAG AT AAT GCA CCC CGG ACC ATA TTT ATT 6343
 Asp Asn Ala Pro Arg Thr Ile Phe Ile
 40 45
 ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC 6391
 Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
 50 55 60
 TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC CCC TGT GAG AAC AAA ATT 6439
 Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile
 65 70 75 80
 ATT TCC TTT AAG GTAAG ACTGAGGCTT ACTTTGTTT CAATCATGTT AATATAATCA 6495
 Ile Ser Phe Lys
 ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA AGTAATGTAA TTAGAATTTCT 6555
 CAAATATCCT CAGACCAACC TTGTTGCTAG AACAGALATA ACAAAGAACCA GAGAACCAATT 6615
 AAAGTGAATA CTTACTAAAA ATTATCAAC TCTTACCTA TTGTGATAAT GATGGTTTT 6675
 CTGAGCCCTGT CACAGGGAA GAGGAGATAC AACACCTGTT TTATGACCTG CATCTCTG 6735
 ACAATCAGTC TTTATACAAA TAATAATGTA CAATACATAT GTGAGTTATA CATTAAAGAA 6795
 TAACATGTCA CTTTCCAGAA TGAGTTCTG TATGAGAAAT GAGCTAATT ATCCTTCTAT 6856
 ATTCTACAC CTTTGTAAAT TATGAGAAAT TTGTTACCTC TAGTTGTTT CTTGCTGATC 6916
 CTTACCTCA GTCTTAGACA CAAGCTTCAG CTTTCAGTTG ATGTTGTTA TTGTTAATG 6976
 TAATCTAAAT GAATAAAAGT TATGAGATCA CCTGTAAGAAC TAATGCTATA ATTATCTTCA 7036
 AGCCAGGTAT AAAGTATTTC TGGCTCTAC TTTTCTCTA TTATTCTCCA TTATTATTCT 7096
 CTATTATTT TCTCTATTTT CTCATTATT GTTAGATAAA CCACATTAA CTATAGCTAC 7156
 AGACTGAGTC AGTARGAGTC GECAGGATG CTTACAAATT GCAATTGCTT CAGAGGAGAA 7216
 TTCCATGTCA TGAGAGACTT TTTTGAGTGG AGATTGCTA ATAATATATCC GCTTTCATGC 7276
 CCACCCAGTC CCACATGAA GACAGTTAGG ATATGACCTT AGTGAAGGTA CCAAGGGGCA 7336
 ACTTGGTAGG GAGAAAAAAAG CCACTCTAAA ATATAATCCA AGTAAGAAACA GTGCTATATGC 7396
 AACAGATACA GCCCCCAGAC AAATGCTCA GCTATCTCC TCCACCCAGA GTGCTACCC 7456
 TTCAAGGTGAC AATTGGAGT CCCCCATTCTA GACCTGACAG GCAAGCTTGT TATCAAAATA 7516
 GCATAAGAGG CCTGGGATGG AAGGCTAGGG TGGAAAGGGT TAAGCTGCT GTACTGAAAC 7576
 AACATATAATTA GAAGGGAAAG AGATGGCCAA GTCAAGGTA TGTGCTGATAG AGGAAAATCTC 7636
 AGTGGAGAG GCAAGATGAG AACACTGGAT AGTCTGGAG CTACAGGTGG ATCTCTGTTG 7696
 AGAGGAGCTG GTGAAATGTT TAAGGAGTC GAAATTAATG TTGCTGCTTA GTAGCTG 7756
 GGAAATATCA TATTTGGGG AGCTGAAAT TTATTCATTA TTGATGCTCC TTGTTAATATAA 7816
 AAAGAATGCTG GCTGGGCTGT GTGGCTCTAC CTCGTAATCC CAGCAATTG GGGGGGGAG 7876
 GGGGGGGGGT CACCTGAAAGT CAGGAGTTCA AGATGAGCT GACCAACATG GAGAAAACCC 7935
 ATCTCTACTA AAAATACAAA ATTAGCTGGC CTGGCTGCTA TATGCTGTTA ATCCCGCTA 7995
 CTGGGAGGC TGAGGGAGGA GAATCTTTC AACGGGGAG GCAAGGTTG CGATGAGGCT 8055
 AGATCTGCTC ATTGCACTCC AGCTCTGCTA ACAAGAGCAA AACTGCTCT CAAAAAAAAA 8116
 AAAAAAAAG TGAAATTAC CAACGCTATT AGCTTAATAA TTAACTACTG TTGTTAAGTA 8176
 GGGGGGGGG TGGCTGGAAG AGATCTGCTG AAATGAGGG AATGCTGCTT TAAGCTTCT 8236
 CAGCATCATA GCAAATCTGC TTCTGAAAG AATTCATTA ATATTAGTTG GAGGGGGGG 8295
 GAGAGTGGAGG GGTGGACTAG GATGAGCTT AGCTGCTGCT TTAACTGCT TTGCTGCT 8356
 CTAATAAGGA TCTTAGCTGT GTGTTATAAA GTGCTGCTAGG TTCTAGATAA TAAATACAA 8416
 CAGGGCAGGC ACAGTGGCTC ATGCTCTATA TCCAGCACT TTGGGAGGGC AAGGGGAGTG 8476
 TCTCACTTGA GATCAGGAGT TCAAGGAGAG CTCGGCCAGC ATGCGATAC TCTGCTCTCA 8536
 CTAAAAAAAA TACAAAATT AGCCAGGCA GGTGGCATGC ACCTGTAATC CCAGCTACTC 8596
 GTGAGGCTGAG GGCAGAAGAA TGGCTTAAA CCAGGAGGTG TAGGCTGAG TGAAGCTGAGA 8656
 TCGCATTACT GCACTCCAGC CTGGGGGACA GAATGAGACT TTGTCCTAAA AAAAGAAAAA 8716

GATACAACAG GCTACCCTTA TGTGCTCACC TTTCACTGTT GATTACTAGC TATAAAGTCC 8776
 TATAAAGTTC TTTGGTCAAG AACCTTGACA ACACATAAGAG GGATTTGCTT TGAGAGGTTA 8836
 CTGTCAGAGT CTGTTTCATA TATATAACATA TACATGTATA TATGTATCTA TATCCAGGCT 8895
 TGGCCAGGGT TCCCTCAGAC TTCCAGTGC ACTTGGGAGA TGTTAGGCTA ATATCAACTT 8956
 TCCCTGATT CAGATTCAAC CCCTTCTGAT GTAAAAAAA GAAAGAAAATE 9016
 CCTTTCCCT TGGAGCACTC AAGTTTCACC AGGTGGGCT TTCCAAAGTTG GGGTTCTCC 9076
 AAGGTCAATTG GGATTGCTT CACATCCATT TGTATGTCAC CTTCCCTATG ATGGCTGGGA 9136
 GTGTCACACA TCAAAACTAG GAAAGCTACT GCCCAAGGAT GTCCTTACCT CTATTCTCAA 9196
 ATGTGCAATA AGTGTGATTA AAGAGATTGC CTGTCTACCC TATCCACACT CTCGCTTCA 9256
 ACTGTAACTT TCTTTTTTC TTTTTCTT TTTTTCTTT TTGAAAC GGAGTCTCGC 9316
 TCTGTCGCC AGGCTAGAGT GCACTGGCAC GATCTCAGCT CACTGCAAGC TCTGCTGCC 9376
 GGGTCAAGC CATTCTCTG CCTCACCCCTC CCAAGCAGCT GGGACTACAG GCGCTGCA 9436
 CCATGCCCCAG CTAATTCTT GTATTTTCTAG TAGAGACGGG GTTTCACCGT GTTACCCAGG 9495
 ATGGTCTCGA TCTCCTGAAC TTGTGATCCG CCGCCTCAG CCTCCCAAAG TGCTGGGATT 9556
 ACAGGCGTGA GCCATCGCAC CGGGCTCAAC TGTAACTTTC TATACTGTT CATCTTCCCC 9616
 TGTAAATGTTA CTAGAGCTT TGAAGTTTG GGTATGGATT ATTTCTCATT TATACTTAAAG 9676
 ATTTCAGATT AATTCACAAAT TGATGCCAC AGCTTAGGCT CTCTCCCAA ATTGTATATT 9736
 GTAGACAGCT GCAAGTGG GTGCCAATAG GGAACTAGT TTATACCTTC ATCAACTTAG 9795
 GACCCACACT TGTTGATAAA GAACAAAGGT CAAGAGTTAT GACTACTGAT TCCACAACTG 9856
 ATTGAGAACT TGGAGATAC CCGTGTACCT CTGTCATCCA GAGTCTTCA GGATCTTTG 9916
 AAGGATGAAAG AAATGCTATT TTAATTCTG AAGTTCTCT ATCACTGCTT AGGATCATGG 9976
 GAATCTGTC TGCCATGAGG CCAAAATTAA GTCCAAAACA TCTACTGGTT CCAGGATTAA 10036
 CATGGAAGAA CCTTAGGTGG TGCCCACATG TTCTGATCCA TCCGCAAAA TAGACATGCT 10096
 GCACTAACAG GAAAAGTGC GGCAGCACTA CCAGTTGGAT AACCTGCAAG ATTATAGTTT 10156
 CAAGTAATCT AACCAATTCT CACAAGGCC TATTCTGTGA CTGAAACATA CAAGAACCTG 10216
 CATTGGCCT TCTAAGGCAG GCGCCAGCCA AGGAGACCAT ATTCAAGGACA GAAATTCAAG 10276
 ACTACTATGG AACTGGAGTG CTTGGCAGGG AGAGACAGAT CAAGGACTGC CAACTGAGCC 10336
 AATACAGCAAG GCTTACACAG GAACCCAGGG CCTAGCCCTA CAACAAATTAT TGGGTCTATT 10396
 CACTGTAAGT TTAATTTCG GGCTCCAGT AAAGAGTAAG CTAAGATTC TGGCACTTTC 10456
 TGTCTCTCTC ATAGTTGGCT CAGAAATGAG AACTGGTCAG GCGAGCATG GTGCTTACA 10516
 CCTGGAATCC CAGCACTTGG GGAGECGCAA GTGGGAGGGT CACTTGAGGC CAGGAGTTCA 10576
 GGACCAAGTT AGGCAACAAA GTGAAATACC CCCTGACCCC TTCTCTACAA AAATAAATT 10636
 TAAAAATTAG CCAAAATGTTG TGGTGTATAC TTACAGTCCC AGCTACTCAG GAGGCTGAGG 10696
 CAGGGGGATT GCTTGAGCCC AGGAATTCAA GCGTGCAGTG AGCTATGATT TCACCACTGC 10756
 ACTTCTGCTC GGGCAACAGA GCGAGACCT GCTCAGAACG AAAAGAAA AGAAACTAGA 10816
 ACTAGGCTAA GTTTGTGGGA GGAGGTTCATC ATCGTCTTTA GCGGTGAATG GTTATTATAG 10876
 AGGACAGAAA TTGACATTAG CCCAAAGGC TTGTGGTCTT TGCTGGAAC CTACTTAATC 10936
 TTGAGCAAT GTGGACACCA CTCATGGGA GAGGAGAGAA GTAAAGCTTT TGATGTATAG 10996
 GGGAAAACCA GAGGCCTGGA ACTGAATATG CATCCCATGA CAGGGAGAAT AGGAGATTG 11056
 GAGTAAAGAA GGAGAGGAGG TCAGTACTGC TGTTCAGAGA TTTTTTTAT GTAACCTTTG 11116
 AGAAGCAAAA CTACTTTGT TCTGTTGGT AATATACTTC AAAACAAACT TCATATATTC 11176
 AAATTGTTCA TGTCCTGAAA TAATTAGGTA ATGTGTTTTT CTCTATAG GAA ATG AAT 11233

Glu Met Asn

85

CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG	11281
Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Glu	
90 95 100	
AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA	11329
Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser	
105 110 115	
TAC GAA GGA TAC TIT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA	11377
Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys	
120 125 130 135	
CTC ATT TTG AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC	11425
Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe	
140 145 150	
ACT GTT CAA AAC GAA GAT TAGCTATTAA AATTCATGC C	11464
Thr Val Gln Asn Glu Asp	
155	

(2) INFORMATION FOR SEQ ID NO:14:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28994 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION: 1..15606
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 15607..15685
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 15686..17056
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 17057..17068
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 17069..20451
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 20452..20468
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 20469..20586
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 20587..21920
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 21921..22054
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 22055..26827
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 26828..27046
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: 3'UTR
- (B) LOCATION: 27047..28994
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ACTTGCCTTA	AAAGCTTTGC	ATAGGTAGAC	AACATTAGAT	TAATTCCTT	GCTCACATCT	60
GTTCAAGAAA	AATCATTAA	GTTATAAAT	ATAACAAACC	TTCTGCATTA	TAAGACTGAT	120
GTTTAGAAAT	ATAAACATTT	TATACATCA	CATTAAATC	TTTCTCCAAG	GCTTCATCTT	180
TATAAAATAG	TCCGGAAATT	TCAGAGAAAG	ATGAATCTGA	TTTTCCAAGA	GAGGACAGCT	240
GTGGACTACT	TGGCACTGGA	GACTAAATAA	AGAAAGCAGG	TACAGTCAT	AAGATCTTCA	300
GGACATATAC	ATTTTGTAA	TTAAGAAAAA	GCAAATAAAA	CATTTTCAG	AAAAAGGCAA	360
ACATGCTAGA	AAGCATATGA	CTTACTGATT	TGAGTTTTA	TTATTAAGGA	AATTACAGG	420
CCCAAGAAAC	ACCTTGCTCA	ATATATTAAA	TTTTATTG	GTGTTCAACT	AGACTTTGCT	480
TTTCATTG	TTGTTTTGT	GA	CAAGTTCT	CGCTCTGTCA	CCTAGGCCAA	540
CACAATCTTA	GCTCACTGTA	GCTCCTAGA	TTCAAGTGAT	CCTCCTGTCT	CAGACTCCTG	600
AGTAGCTAGG	ACTACAGGAA	CATTGACCA	TGCCAGCTA	ATTTGTTTT	GTTTGTTTT	660
GTTTCAGAG	ACAATGTATT	GCAGGTTG	CCAGGCTGAT	CTGAAACTCT	TAGCTCTAAA	720
CGATACTCCT	GCTCAAGCT	CCCAAAGCAC	TAGGATTACA	GACATGAGCC	AATGCGCCCA	780
GCCTTAAATT	AGACTTTAAA	TGTGTTTA	AACTCCTGTT	AAAAAAGCGT	CTGCTATCTT	840
GAACCAGTAG	ATGTTTCTAT	AGCAATGAAG	CTAAACTGTA	ATTTAGACAG	TAGCCAAATG	900
CTTGTGAAAT	TTTGTAAAT	AATATATCT	TCAAGGGAGC	AAATCATGTC	CCAAATGCAA	960
AAGATCAACT	GTTGGGGCA	GTAGAAAAAG	ACAGGATACT	GTGCTCTTA	AAAGGTCAGT	1020
AACTATAGTA	CCTAGTTATC	TACTTATCA	CAGCAAAATG	ATTACATAAA	ATCCTATGGA	1080
TCATAAAGG	ACAGACTCAC	TTCTGTCTCT	AGATCTCAAG	CTACCAAAAA	GAAATCTCCC	1140
AATAGTTCT	TGGAGGCTA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTTCCCTCC	1200
TATGTTGTAG	GACAATCTA	GCTCTGGCA	TACGAATA	TTAAATCCCA	CTTATCTATA	1260

GAGCTTTCTT AAAGGGAAAGA AATTTGAGTA GTATGTAAAAA CAGAATAAAA GATTAAGGCT 1320
 CCATAGGCAT ACAGCTTACG TCCAATTCTC TTGGCCTCTT GCAATTCTA TTATCAGGCT 1380
 TTACAAGGTG ATTTGJCATC ATATTCCGAA GGCACCAAGCT ACAAAAGCTTA GAACAATGCC 1440
 AGATTTAGCT ACAAAATCTCA TGCTACAAGC TCTCTGGAAT CCTTCCTGT TTOCCACTCC 1500
 TACTGCTGAT GTTAATTCTAG ACTGTCTATTA TCTGTCACTT TCTTAAACTC AATTCTCTC 1560
 TCCCTTAAAT CATTCTATCA ACTGCTATTG GGCTAATCTT TCAAAAACCTT GATTACTGCA 1620
 TTCTTAAAC TCAAAAACCTT TCATTGTTCT AGAATAAGCTT GAAAATCTCAT GATATGGCT 1680
 TCAAGGTCT GTTATTCTCG GTGCAAGCT ACTAGTCCCA TCAATTCTCAA CTATCCTCT 1740
 CTATGTACTT AGCCAAATGCA GTCTCTCTGG CAATTCTGCC TTGTTTCAGG ACTGGCTCAG 1800
 TTAAGATTCT TTTATCTTCG GCGGGCGCG CTGGCTCAAG GGTGTAATCC CAGCACTTTG 1860
 GGAAGCTGAG GCAGGAAGAT CACCTGAGGT CGGGAGTTCG AGACCAGCCT GGCAGCATG 1920
 GTGAAACCT GTGTCTACTA AAAATCCAAA CATTAGCCAG GCGTGGTGG AGGGCGCTGT 1980
 AATCCCAGCT ACTTGGAAG CTGAGGTGAG AGAATCGCTT GAACCCAGGA GAGGGAGGTT 2040
 GCACTGAGCC GAGATTGTGC CATTGCACTC CAGCTGCGGC AACAGAGGGA GACTCCACCT 2100
 CAAAAAAGA AAGGATTCTT CTATCTTCAC AATTTCTTAA TGTCTTAAACA GGTCTTACAG 2160
 TTCAATCTAAT TCAATCTCAT TTTTACAG TGAGAAAACA GGGACAGTGA CGGTGGATCA 2220
 AGTGACACCA GTAAAGACTGA GCTAAATTAG AAGCGAGATC TCACTCGAGT CTGAGGTTAT 2280
 TCCCACTGTC CAACCTTAAT TTAAAGTAGC TTCAAAATTCTT ACTTTTACTT TTCCATAAAAT 2340
 TCGGAAGGAA TTTTCCTAG GAGTCAAAAT GTTGAACCTT GGAAGGGTAT AGTCTCTGTG 2400
 TCTTGAGAT GAGGGAGCC CTGCTCATAT TCAAGTTCTT AATTGACTTT GTTGTGTTTG 2460
 AGAAACGATG CTGATTGGG TAACTTTAAC ACATCTGTTT GATTAGTCTT ATAAAATATG 2520
 CATATATAGA AGACAGAAAAG AGCAACAAACA AATTTGAAAG GTATCGTACG TGTCCATTCC 2580
 GTATCGTACG TGTCCATTCC TGCCAGTACG TTTATAGTAT AGTATTAAATA GTATCTAGAA AATACTACAC ATGCACAGCA 2640
 GTTGGAAAAT ACTTCAGAGA AGCCAACAGG CAGATTTC TCTCTTCCCT TCCCTTCTA 2700
 ATTTTCCCTT TCCCTTCACT CCCCCTCTCT TCTCTCCCA AGTAACACTG TGCACTATG 2760
 TCAAAACGAAA ACTTATAATC AAGTAATCTGTTT TTCTGCAAAAT TAAAGTFOF TTTCTGTCA 2820
 TGGETCAAGG CCTCAGCAGA TCCAGGCTG GGGGAGGGC TGGTCTTCGT CGTGTGCCAA 2880
 ACACGTACCA CTGCTCTGGC TCTGCACTT TAGGCTTAGT GAACGTGGCTG TTACTAAGCA 2940
 CTGCTCCCTC TGCCCTCATG AGCTGCTCTC TTCTAGTCTT CTCCTCTTC TCAACGCGAT 3000
 CCTAGCCCTC CAGGCCATTCT CACCTCTTCTT TCTCTCACT CCCCCTGCG CTCCTCGACT 3060
 TCTCTCTAC TGTGTTCTCC GCCCCACTAG AGCCCCCTCAG AGAAAAGTTC CTCCTCGCA 3120
 CCTTCTCTG TGTCACAGCC CTCACATCT TCAAGGGCGC CGATCCCTCC AGCCCCAACCC 3180
 CAAGGCCAAT GTACTCTGGG GTATGGGAC CTTCTCTGTC AGCGAACCGCG AGGGAGTGA 3240
 GACCGTGGG GCGGGGTGCT CGGACTTCG GGGGGAGGT GGGAAAGGCGC CGGCACTCC 3300
 AGCAAGCCCT GCAAGAGTCG CGTGCAGAGCT CTCCACCCAC CACCCCCCCC AATTTCCCGA 3360
 CGCTAGCTC CCAGAGCCAG GCCCCAGGGA AAGGAGCTT TTTCCCGTT TTCTCCCGCT 3420
 CTTTCCCTC CACTTGAAT ACTCTGAAA CAAATACTC TCCCTGCA CCGTGTGTG 3480
 TTTGAACCGA GAAAAAACTCT GAAACTGGTC AAGGAAAGAAC AAGGAAGACT TGCCAAAGCA 3540
 AGGCCGGTGT GTGTCCCAGC AGCTTAAAT CTCAAGCAAG GAACATAAAA TAGCACATTC 3600
 AGGGCCCTCTT TTGAGTAAATTTAATTAATCTG TTTGTTTCTCA GGAAGGTTT AAAACTGCT 3660
 TTGAGATG TCTGTTCTCA GSAASCTTCTT AATCAAGCTG TCTCTGCG CACAAGCAG 3720
 GTTTTAGAT CCAGAGCTC AGTTAATGCG CCTCTTCTC TCTGGTGTG TCAATAAGTT 3780
 CAGAATCAGG CCTCTTCTAGA AATTTCTAC CTCGGGTGTG CATATAATTG TAGCTTTGDC 3840
 GCAACAGAT ATCAAAAAGT GTGCTAAAC ACACATGGCT 3900
 TCATGGGTG TTTTAATGCG GAGGGTTTGA CTCGAATTG CAAAGATATA CATTCAAGC 3960
 TTACGCCAG TTAGTGGATG TGGAAAGAAA AAAAAAGCAA ATTACCTCAT AACACAAAGC 4020
 TCAATAACAC ACATCCATAA GTCAGGAGTA CAAATACTTAA CATCTTAGAG ARCTATATT 4080
 AACATTTACA TACATTAATC AGGTCTTTTT TTCTCTTCTC CITGATTAATG TGTAGTTAT 4140
 CATTAACTCT TGGATTATTCT CTGCTGTGTT ATATTCTTATTT GGTGTTGTG AAGAGGCCG 4200
 TTGTTTAAATA TAAGTCTCTA GAAATCTG GTCATCTG TTAATCTGA GTGCTAAATA 4260
 TTGIGAAATA TAGGCCACAT AATACTGCG TAGATAACTA TGGCAAGAGTA AGGAGTCTCA 4320
 AACACTGTCG CAGAACATA GCAATCTG TGAAATTCTT ACCTCTGTC CTAAATGAA 4380
 GGGAAAAGGA ATGAGTTTCTT AGTTTCTCTT AATTTCTC TTTATGTTT CAGACTCTC 4440
 AGCAAGTATAA AGTTTCTAC AAGTAAATA TATTCCTT AAAGTGAATG TGCTTTATTC 4500
 TGATACCATG TCCTTCTAA TTGGGGGGC CAGGTGAGAT AAGTTTATG AAATAAAAAG 4560
 ATAAAAAAAT CTTACATTCT TAGTGTCTT CTTGGTAAAT ATGAGATGTT GTCCACTGTG 4620
 TTATCTCTT CCTCCTTATT ATGATGGTGT CTCTTATTAT TTTAATGTT TCAATTAAAC 4680
 CAAGGCTG GAAATACTC ATGCAATTCA TCTCACAGCC TTCACTCTG ATGATATTAA 4740
 AACAGGTGATG TGTCCATCTG ATTCTTAAATA TATTCCTAAG AAAATGAGT CTCACCTAATG 4800
 CATAAATGCT TTCACTCATGAT TAAGAGAAACA CTCATGGACAT TTIATTCTTAT TTATTTTTT 4860
 AAATATTAAC TTCCATTGCA TAAGCTAAAT GGGTAGGAAT AAGTGAATG ATATTGTTAT 4920
 CTAGAGCTTT AAAATATTCA AAGGGCTGTC ATCATTATCT CATTAACTCT TTGAAAACAA 4980
 CTCTATGAG TACAAAGGAC ACTGAGACAT TTGTGCTCT ATATCAAAGA AAAAGTGT 5040
 TGCCCCAAA CTTCAAAATG TGAAATTAC ACATCTGCA TCITTAACAGC TGAGAAAAT 5100
 TCACTGGCAA TGGAAATTCTT AAAATTAGAG CTTGCTTAGT GTGCTGCTTC TGATCACTAC 5160
 TTGATCCAC TTCGTGCTT CATGTTAATT GGCCTAATTG GACTCTACAG TTGGAAGGTG 5220
 5280

AAAACCTTACT	ATTTCAACTT	GAGTCACGTA	TGTATTCTTA	TATATACTT	CTTAAAGGTA	5340
CTATTTTTTT	TCTTCTGATA	GTCAACACAC	CAAGCACTTC	CAGCCACCC	GTCACAGACT	5400
TCCTTTGTAA	TCACTGTTGA	AGGACATGAT	GTTTTTATGA	CTTCCCAGAAA	TGAAACACCT	5460
ATCTTGTTTT	AAAAACAAAC	AAACCAACAA	AAAGTAACT	TTATGTAAGC	ATTTGTTTC	5520
CTGACTCTAG	GAACCCCT	GTTTTTATAT	CAACTCTGTA	CTGGCAAAAC	ACAAAAACAA	5580
AATGCCACT	TGCTAATTCC	CTTCTCTAGTA	AAGTAATACA	GTTAGCACA	TGTTCAAGAA	5640
AAAAATGGCT	AGAAAATT	GTTCACATA	ATTATTTCA	AGACTGTGAT	ATTTACACTC	5700
TGCTCTCAA	ACGTTACATT	TTATAAGACT	ATTTTTAAC	ATGTTGAACA	TAAGCCCTAA	5760
ATATATGTAT	CCTTAATTG	TATTTCAAAAT	ATTTTAGGTC	AGTCTTTGCT	ATCATTCAG	5820
GAATAGAAAG	TTTTAACACT	GGAAACTGCA	AGTAATATT	TGCCCTTTA	CCTGAATT	5880
GGTAGCCCTC	TCCCCAAGCT	TATTTCTCT	TGCAGAAAAGT	GTAAAATTAA	TTACATAAAA	5940
TTCTAATGAT	GSTATCGTG	TGGCTTGAT	CTGATACAGC	AGATAAAGAA	GTTTTATGAA	6000
AATGGACTCC	TGTTCCACTG	AAAAGTAAAT	CTTAATGCCC	TSTATCAACT	ATCTTTGAC	6060
ACCATATTGA	GCTTGGGAGG	AAGGGGAAGT	CCTGAATGAG	GTTATAAAGT	AAAAGAAAAT	6120
ATTGCAAAA	TGTTCCCTT	TTTAAATGT	TACATTTAG	AAATATTTTA	AGTGTGTA	6180
CATTGAGGA	ATTACCCAA	TAGGACTGAT	TATTCGGAT	TGTAAATTA	GAAAAGTTT	6240
TGTGCTGAAG	TGTGACCGG	AAGTCTGAAA	ATGAAGAGAG	ACAGATGACA	AAAGAAGATG	6300
CTTCTAATGG	ACTAAGGAGG	TGCTTCTTA	AASTGAGAGA	GAGATACTCA	GAAAGAGGTA	6360
CAGTTTTG	AAGGCACAGA	GCCCCAACTT	TTAAGGAAAGA	AAAGATTCTA	TGAAGATAGT	6420
GATATTACAT	TAAAAGAAGT	ACTCGTATCC	TCTGCCACTT	TATTCGACT	TCCATTGCC	6480
TAGGAAAGAG	CCTGTTGAA	GCGGGGCCA	AGGAGTCCCG	ACAGCAGTCT	CCTCCCTCCA	6540
CCTCTTCCT	CATTCTCTCC	CCAGCTTCT	GAGGCCCTTG	CTCCCCCTGGC	GACTGCTGG	6600
ACAGTCAGCA	AGGAATTGTC	TCCCCAGTGC	TTTGCCCTC	CTGGCTGCCA	ACTCTGGCTG	6660
CTAAAGCGGC	TGCCACCTGC	TGCACTCTAC	ACAGCTTCGG	GAAGAGGAAA	GGAACCTCAG	6720
ACCTTCCAGA	TCGCTTCCCT	TCGCAACAAA	CTATTTGTG	CAGGTAAGAA	ATATCAATTCC	6780
TCTTTATTTG	GAAAGTCAGC	CATGGCAATT	AGAGGTAAAT	AAGCTAGAAA	GCAATTGAGA	6840
GGAAATATAAA	CCATCTAGCA	TCACTAAGAT	GAGCACTGAG	TATCAACATA	AGAAATATAA	6900
GCAAAGTCAG	AGTAGAATT	TTTCTTTTA	TCAGATATGG	GAGAGTATCA	CTTTAGAGGA	6960
GAGGTTCTCA	AACTTTTGTC	TCTCATSTTC	CCTTACACT	AAAGCACATCA	CATGTTAGCA	7020
TAAGTAACAT	TTTTAATTAA	AAATAACTAT	GTATTTTTT	AAACACAAAA	AAAAGATAAA	7080
AGAGTGACAA	TTTTTTATTT	TTACAAGTGT	TTTAAGTGGT	TTAATAGAG	CCATATAGAT	7140
CTGCTGGATT	CTCATCTGCT	TTGCATTGAG	ACTACTGCAA	TATTGACAG	ATGCAAGCCT	7200
CTGTTAAACT	CTGTTGTACA	CTCATGAGAG	AATGGGTAA	AAAGACAAAT	TACGCTTTAG	7260
AATTATTAGA	AATAGCTTC	ACTTTAGAA	CTCCTTGAGA	ATTGCTGCTT	TAGAGTGGTA	7320
AGATAAATAA	GCTTCTCTT	AAACGGAAATC	TCAAGACAGA	ATCAGTTACA	TTAAAGGCAA	7380
ACAAAAAAATT	TGCCCATGCT	TAGTCATTTT	GTGAATCTG	CCACACCTT	GCACTGGGCT	7440
ACAATTGGAT	AATATAGCAT	TCCCCGAGAT	AATTCTCTT	CACATTAAAG	GAAGGGGTG	7500
AATAAAATAC	TCTGTTGAA	GTTGAATTAAC	AAAATTTAGG	ACCCCTAAA	TTTTAGGCT	7560
CCTGAAATTC	GTCTTTCTC	CTATATTGAG	CTACCTTACG	TTTATTAATA	TCTTTTCA	7620
GGCAGGTTG	ACTAGCTCAT	GGCTTACAGTC	TCAGGCCAGG	CTAGGCCAG	GAATTTGAGA	7680
CCAGCCAGGG	CAACACAGTC	TCTACAAAAA	AATAAAAT	TACCTGGGTG	TGTTGGTGCA	7740
TGCTGTAGA	ACTACTCAGG	ATGCTGAGGA	CTGTTTGAGG	CCAGGATAG	CAAATCTGTC	7800
GTGAGTTCTAG	CCACTAAACA	GAGCGAGACT	TTCTGAGAAA	AAATAACAAA	AAACAAACAA	7860
AACTTCCTTC	AAAATAACTT	TTTATCTGCA	ATGTTTTCT	ATTCGCTGTC	AGATTAATT	7920
TACTCTTTTA	CCTGATTCTC	AAAGGCTTC	ATATCTATT	CGAGCTTTAC	CTTGGTGTCA	7980
CTGCAAAATA	GCAGGACTGT	TCCACTACAA	TCCAAATTC	ADAGGTGGG	TGCACTGGCT	8040
CACTCTGTA	ATCCCCAACAC	TTTGGAGG	CAAGCGAGGT	GGATTGCTTC	ACTTCAGGAG	8100
TTCAAGACCA	GCCTGGGCAA	CATGCGAAAA	ACCTCTCTC	TCTAAACAT	ACAAAAAATTAA	8160
GGCAGATGTC	GTAGTATGTC	CCTGTAATG	CAACTACTTA	AAAAGCTAAG	GTAAAGGGAT	8220
CACTTGAGG	CAGGAGGTC	AGGCTACAGT	GAGGCACTGTT	TACTGTGTC	CTGCACCTCA	8280
GGCTTCTGAA	TAGAGCAABA	CTATCTCA	AAAGGAAATA	AAAGGAAAGA	AAAGGAAAGAA	8340
ACATGCGTC	ATTCAGTTCA	CCCCACAA	AAACATTGTT	TGATTATCAC	ATAAAATGTC	8400
GTCGATTGCG	TTCTCTATCT	ATTCACATT	TTAACGATT	TGAGGATTC	AACTCAATT	8460
TCTTTCTCAA	ACTAGGCTAT	TTAACACTAA	TGAGTTGAT	TTTGTATTTC	TTGTTTTGAG	8520
TCTACAGATT	CAAAACAAA	AACTAAAAAA	CTTATTTT	AAATTTCTG	CTACGTCAC	8580
TTCTTCAAGA	CTCACATACA	CACATTATA	ATAAATGCG	AGAATGTC	AGGATAAAAT	8640
GATTATATA	ACTGAAAAT	TAGTTTTGAG	TCTTCTTCT	GTCAAGATG	CTACCTACCT	8700
GATCTCAAGT	AAATTAAATT	GTAGCATGTC	CCCTCATTC	ATCCCATAC	TATTCACAG	8760
GATTGGAATT	CCACAGCAAG	GATAAAACATA	ATCTAATTG	CTTTTCAGT	TCAAGGTATT	8820
TTAACCTTIA	ATCTAGTAACT	ATGTTTGTG	TGTTGTT	TGTTGAGAT	GGAGGCTTG	8880
TGCTGCACTCC	AGGCTGGAGT	GGAGTGGGAA	GAACCTGGGT	CACTGCAACC	TGTGCTCAT	8940
GGGTTCAATTC	AGTTATTCTG	CCTCAGTGTG	CCAACTAGT	GGGACTACAA	GGCACATGCC	9000
ACCATGCGTC	GCTAATTCTT	GTATTTTAA	TAGAGGAGG	GTTTCACCAT	GTTGGCCAGG	9060
CTGGTCTCGA	ACTCCGTGACC	TCAAGTGTG	CAGCCGCGTC	GGCTCTCCAA	AGTGTGGGGA	9120
TTACAGGCT	AAGCCACCGT	GGCCAGCTA	ATAGTATGTT	TTTAAACTCT	TAGTGGCTTA	9180
ACAATGCTGG	TTGTATATA	AATATGCGAT	AAATATTAA	TGTCTTAGAA	TTATGAAGAA	9240
GTGGTTACTA	GGCCGTTCTG	CACATATCAA	TGTTCTCTC	CTTACAGCTT	TAATTAGAGT	9300

CTAGAATTGC AGGTTGGTAG AGCTGGAAACA GACCTTAAAG ATTGAATGCA CAACTTCGTT	9350
GTCCTAAATGA GGGAACTGAG ACCCTTAAAAA TTAAGTGACT TGCCCTAGAC AAAACTGAA	9420
CTCATGTGTC CTAATTCCA TCAATGAAATT CTACCATTCG CTACGCTCTG GCTAGTTGTC	9420
AAAGTATTGC ATAACATAAT TTTATGTCT GTTTTAAAGA ACAAAATTGTC ACTGCTTACT	9540
CCTGGGAGGG TCTTTCTGAG GTGGGTTATA ACTCTTAAAA AAAAAGAGATCAGTAGTCTG	9600
AGAATTCTAG ACGAAATAGT CAAAGCATT TTATCCATG GATCTATAAT TTTCATAGAT	9660
TAGAATTAA TCAAAAGAAAC ACGGATGAGA AAGGAGAGG AAAATTGAGG AGAGGAGGAA	9720
TGGGGATGAG AACACACTAC TTGTAATCAG TCATAGATGT ACTGAAACT AACAGAGA	9780
ATTGTAAGAA AATAAAATG AAAAATTCAA AACCAACACA TGAAATAAAA AGAAACTACT	9840
AGGGAAAGAT GAGAGAGACA TTAGAAAAAT TATTCTATTG TTAAGATCT GTTTTCAGGC	9900
TTCCCTCTG TTCTCTCC TTCTCATGGG TTTTCAGGTG GAGGAAAGT TTAAGATGGA	9960
AAAATATATAT ATATTCTACA CATECCTTC TAGCTGTTG TCATGCAAC AAGTTTATC	10020
ATAGCCTAACT TTTATTCTATA CAAACATTAT TGAGTTCTTA CTGCTGCTA AGCTCTTCC	10080
AGGTGTTGAA AATTCAAGGG AAAAAGACA ACTCATTGTC TTAAACTCA GATGAAAGCT	10140
GAACAGACCT ATTTTAATC AAAGTAATCT CAATTAGGG TAGTAAGAGC TATTTAGAA	10200
GCATGAACAG GTGTGAAGGA GTAGGACTC TGAGGAGAGA ATAGTAGT AGGAATGAA	10260
GAGCAGAGAA GTTTCTTAG AGGAACATT AAAGCTGGG GTTACGGGAT GAAAGATGAG	10320
GCAGGGTTG CAGGCAAAAA AAAAAGAG GCAGGGGAAG GGGAGTCT GGCCTGGCAG	10380
AGAGAATAAC TGTGGCAACA ATGGAGAGGA GTCTGGAGC AAGAAACCA AGTAGAAGAG	10440
TATTAATAA GAAGATGCCA GGGTAATGAG GGGCTTGATT TAATCAGTG CTGTTGGAGA	10500
TGGAGGAGA ATACCAAATT CTGGAGACAT TTCTGAGTTA GAACTACAG TATTTATCAG	10560
ACAAGGGAAA GATTAGACAA AGGAGTTAG AATGACTCCC AGGTTCTAGT TTGGGGAGG	10620
TAACTAGGAC ATGTTTGAA AAGTAATGTA TTGGATCTCT TACCATGGG ACTATGTATG	10680
TGGAGCCAAA TTAAATTG TACATGTATA TAATCCTCCC CCCACACCA GTAACACTT	10740
CCCTAACTCT CTAACTTGTA GCCAGACTC CTAAAGAAAT AGTTTATG CACTGCTTT	10800
ACTTTTCCCC TCCCATTCTG TCCCTAGATAT TTGTCACCT ACCATCTGCT GCCTCCACTT	10860
TACCCAACT GTTCTACGGT TGCCCAAAAC TTCTTAATTG CCATTCATCA TGACAAAGT	10920
TAAGCTATA TGTAATTAG GAGCTCTACA GTTTGATTG GAGCAGCCCC TTCTAAACC	10980
CTTCTCTTT CGAATTCTGT GACACATCAG AGATTACAA AACGAACTA ATTATTTAC	11040
ACTTGAGCTG TATTTTCGTT CTTCTTTCTT GATGAATGAG GTAAACACTC AAGAAATTGC	11100
CCAAGCCAAA AACTACGAAG TCATCCTCA TTCCCTCTC TTCTGTTGA CCCACAAACAG	11160
ATCAGCTGAG AAATCCCCT GTTTAGTATC TCTTGAATTG ATTACCTAA TTATAGCCT	11220
CATCAACTCT TAATTGTTAA AATTACTTCA GTAGTTGTTG TCTGATCT GTCCAATCTT	11280
GTTCAATCAG GTCCATTCTT TTGTTCTTGG TGTTGGTGT GTGTTGAGA GAGTTTCGCT	11340
TTTGTGCCCC AGGCTGAAGT GCAGTGGAGG ACTTCACTG CAAACAGGC TCCTGGGTTT	11400
AAGCAGTTCA CCCTCCCGAG TAGCTGGGAG TACGGGTATG TGCGAGGACA CCAGCTAAAT	11460
TTTGTGTTTG CAGTAGAGAC AGGGTTTCAAC CATGTTGGC AGGCTGCTC CAACCTCTG	11520
ACCTCAAGCA ATCCACCCAC CTCAGGCTCC CAAAGTGCTG GATTAAAGG CATGAGCCAC	11580
TGACACATGCA CGAGATCCAT TGTTTATGTT GTTCTAGAG TGAGTTTTTA AACACAAAT	11640
TTGACCATAT CTTTCTCCAA TTAAACTCAG TATTTTTTTT TTCAAGGAA AACAGTTCAA	11700
ACTCTTIACT CTGCTTACAC AAGGCCTTTC TAGCTGACT CTTCTTCTCA AGCTTTCATC	11760
AAAGTATAACT GAAAGTTACA TTTTATGAG ATTGAATTAG GCAAGCTAT AACGAAATTAA	11820
GTTTATATG GCAAAATGGA AATAATGTTA ACTCTTCCAA ATAGTTTATC TAGAATGATA	11880
TAATTCTAAA GCTGTCAGGT CAAATGAGTT ATAACACTGTT AACACTATTG CCACATGCAA	11940
GTGTCTCTTA TACTGGTAG ATTATCTG TTUCATGTCA TTATTATGTA AATTAGACTT	12000
TAATAACTC AGAAGTTCTT CAGACATACA GTTTATTATT GTGTTTTTA AACATAATT	12060
TAAATAATT TATATATGAT AATGTTATCC AAATGCTAAAG GGATGTTATG TTACTGCTGT	12120
GCAAAAGAA AAAAAGAA AACTCCAAAT AAATATGTTG AAACCAAGTT TATATGAG	12180
AAGAACATAT TAAAAGGCC AAGTGACAC CATAATGGC TGTGTGAGA CGGCAGGCTA	12240
CAAAACATTA GAAATAATGC TGAGAAAAGT GAAAGAAAC AGAAAGAAAC ATATATGCTTT	12300
GTTTGTGTTGTA GTTTATGTA CTCCAAAGAT ATCTCTCTG AACATTCTAC GTTTTTGTA	12360
AGAAAATGTT AACTTGGCT GGCGGAGTG GCTCTGCT GTACTGCTC GTTTGGGAG	12420
GCCAAAGGGG CGAGATCA TGAGGTCAG AGTTTGAGAG CAGCTGAGC AAAAATGGAG	12480
AAGACCGGCT CCCTCACTAC TAAAGAAATA CAAATTAGG CGGGTGACAG TGCGTTACCC	12540
CTGTGATGCC AGCACTTGG GAGGCGAGG CAGGAAGATC ACCTGAGTC AGGAGTTGA	12600
GACCAAGGAGT GGAGAAACCC GTCTCTACTTA AAAATACAAA ATTAGGGGG GTGGGGGTG	12660
CATGACTGTA ATCCACAGCTA CTCAGGAGG CTAAGGAGAG AATCACTTGA AACCCAGGAG	12720
TGAGGTTTC AGTGAGGCCA GATCTGCTCA TTGCACTCCG GCTTGTCAA CAAGAGGAA	12780
ACTCTGCTAT CAAAAAAACAA AAGAAAGAA AGGTAAACT TGAACATATGT GAGATCTTTA	12840
GAAATGCTT CTTTGTAA AATGCTATTA CATTGCTT ATTATGCTA AAAATGTTA	12900
GGCTCAAC AACCCATATT TTCTGGCTC CCCCCGCTGC TAGCTTGTGTCACATTCTC	12960
TCTTCITGFT GGAAGCTCTT CCTCTGCTG TGAAATGCT TGTTCTCTT TCAAGGTTAGC	13020
ACAGTCATCA CTTTGTGG TAACCTTCTC CAGGACCATC AACAGAAAG ARTGAATCTC	13080
TTGTAATTC AGCTCTTACG TCATTCAATTA CATTATTTG TAACTCTTA TAGATCTTC	13140
TCTCCCACTA GACTCTGAGT CACTGGAGAG TAGGAGCCAA CTCTTATTCA TGTGTGGTTT	13200
GGTCAGCTAC TGGCCACATT CCTGATGCT AGTTAATGCT CAAATCTAA CTGGTGAATC	13260
AGCTCAAATA TTGTCTTCT CTAATCCAT TCACTCATTG ACTAACTATG TACTCAAAAT	13320

AGTAAACACC	AGTAATTAA	TCCAATTCCCT	GCCCATACTG	CTTGGTACAT	TTCAGGTGAA	13380
TTAGTTGAT	AAATATGTGT	GTATTACATA	ATATTAAGT	ATGTACAGAA	GATCATGCTA	13440
ATCATAATTC	AJAACGTATA	ACTAATCAA	CATAAATGCT	CTCAGGTAA	CAAATGTC	13500
CCTCTCTAGT	TAATGCACTC	ATTAACAAAC	ACCTTCTGAT	GCTGATAATA	GGCCCTTGT	13560
CAGCAATGAA	GCCATAAAGG	TGAATAAAGA	ACATGCCCTC	GTGGAGCTCA	CAGCCTAGTC	13620
ATTATTGTTG	TGATTTTAA	TATTAATGTT	GGTTGGGTT	TTGGTGA	ATTTTAGAC	13680
TTATCTTAGT	GATCTTTCA	TCCTTGTCA	TATTATT	CTCTAAGAGT	CTTCCTTATC	13740
CCCTCCTTTA	AAAAACTAGG	TGATAATTCT	AAATTGAAA	TTTAAATATT	ATAAATAGCT	13800
TATAAAATT	AATATTATA	ATATTAAT	GTGGATAAA	TATTTAAAT	TTATAATATT	13860
TAAATGTTA	TTAAATTCA	TTTGTACATC	AGTTTTATT	TTATTTAAAT	GTGGTGGCCA	13920
GGCATGGTGG	CTGACACCTA	TAATCCCAGA	ACTTTGAGAG	GCCAAGTCAG	GCAAACCATT	13980
TGAGCTCAGG	AGTTTGAGAC	CACCCCTGGG	AAAGTGGTGA	AAACCTGCT	CTACCAAACA	14040
TATGAAAAC	TATCTGGTG	TGGTGGCAGG	CATCTGTGTT	CCAGATGGG	AGTCCCAGGC	14100
TAAGATGGGA	GAATCGCTTG	AACCCAGGTG	AGAGGGGTG	GGTGGATGTT	GCACTGAGCT	14160
GAGATGTTG	CACTGCAJTC	CAACCTGGGT	GAAGAGGTGA	GACTCCATCT	CAAAAAAAA	14220
AAATGTTATC	TAAATAAGAT	AAATTTAATA	ACTGTTGCA	CTTAGATGAG	CATAAGGAAC	14280
TAAACCTAGA	AAAATACATC	AAATAAGGGC	TGGGTACAGT	GACTCATGCC	TGTAATCTCA	14340
AGCACTTGG	GAGGCCAAA	TTATACAAAG	TTAGTTGTT	AAACACCAACT	AAACACTATT	14400
TTGGGGTTAG	CTTAATTCA	ATTAATT	TTTAAACTGA	STTTAAATT	CCTGCTTACT	14460
CTACCATACA	TGCTAGGCT	CATATTATG	TAGAAAAATT	TTGAGCAGAG	ATTATGAAT	14520
ACTCTCCTG	ATACCATT	ATTTTAAAC	AAATTAAAT	GCAGTATATA	TGTGCCTTT	14580
TACCAACACA	TTAAATAATA	AGATCTACTG	TGAGGACTAA	ATTTCTGAA	TTTCAAAGTA	14640
GTAATGAGTT	AAACCATGT	CTCAAGATCT	CTGAAATAAC	TGAGCACAA	CAGAAAATAG	14700
GTATTTCTAT	TAATGACAGA	GTCACAAGTA	CTACTAATAA	TACTGTGTT	TGTTTCCTGC	14760
AACTAATCAT	GGGAGGAATG	CTAAATTCA	GAAGTTGCTG	AAATACATG	TGTATT	14820
TCCCGATCCA	AGTTTACAGA	TTTCTACAC	TGAGAACTCC	TATTGATATA	CAAATTCTG	14880
GAAGCCTGCA	CACCGTATTG	GAAGAAGGG	AGAAAGGAAA	AGCAGATGGA	AGGATTAAA	14940
TTTTTTCAA	ATCCTGTATC	CCTTGATT	ACAGCAAGAT	TGTATT	TGTTACTGT	15000
GTAAAAAAAT	TAGTATAATC	GAAGCTCAG	ATCAGGAAAT	ACCGCAGTC	AGGGAGAAA	15060
AGGGCCACCA	AATGCGAGAG	CCCTTCAGG	TTCTCCACG	CTGCGTGTAC	CCTCAGATG	15120
AAGCACTTTT	TTATCATG	TTCACCTTA	GCATTGTTAC	AAATGAGTCA	CAAACCTTCA	15180
GCCTCTCACC	CATAGGAACC	CACTGGTTG	AAAGAGAAGGA	TGAAGCAGT	CCTCCTA	15240
GGGCACGATT	AGATGTGTT	ATGGCATCCT	CAGETGAAAC	TATATTTATA	TTGACAATAT	15300
ATTATATATT	CTCAAGGAAT	ACTAGAATAA	TGATTGAGT	CAGTACTAGG	CCATTATATCT	15360
ACCCCTTATA	ATATTGTTA	ATGAGAAAAT	GCTTCTATC	TTCAAATAT	CTGATGATT	15420
GTAAGAGAAC	ACTTAAACAT	GCTTATTCAT	AAAGTGAAC	TTCTGGCATT	TATGAAATGT	15480
CAAGATTGTT	CATCACTATA	CTAGGTGATT	AACTGACCA	TGAACTTGAA	GTTAGTATAA	15540
AGTAGTAGTA	AAAGGTACAA	TCATTGCTC	TTAACAGATG	GCTCTTGT	TTCATTAGGA	15600

ATAAAG ATG GCT GCT GAA GCA GTA GAA GAC AAT TGC ATG AAC TTT GTG GCA 15651
 Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala

-35

-30

-25

ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAAGGC TAATGCCATA 15702

Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala

-20

-15

-10

GAACAAATAAC	CAGGTTCAGA	AAATATGTT	CAATTAGAAA	AGATGTGTTG	AGGTGAAC	15762
TTAAGTGA	CTTTGTGTC	CCAAATTICA	CTCTAATATT	AAATGTTCTT	AAAAAAATAG	15822
TGGAACCTCTA	GAATTAAAC	AAACACATGTC	CAAGGTGTC	GCACGTTGTC	ACACCCACGTG	15882
TCCTGGTACT	TTAATCAGCA	GTAGCTCACT	CTCTGAGTGG	CAGTAAGTGC	ACATCATGAA	15942
AATCCCAGTT	TTCATGGGAA	AAATCCAGTT	TTCTATTGGA	TTCCATGGGA	AAAATCCCAG	16002
TACAAAATG	GGTGTCTCA	GGAAAATACAA	TTTCGCAAG	CAAGATGGA	AAATTATGTA	16062
GAGATTCTG	AAATTTAGAG	TTCGTTGAAT	TACACATT	TATGAAATA	TGTTTGACAA	16122
GTAAAAAAATG	ATTGTTTTT	TTTCTTCTG	TTTCTTCTG	TGGGATGTC	TGCGACAA	16182
TCTGCTCACT	GCAACCTCTA	CCTCTGSGT	TCAAGCAATT	CTCTGCTC	AGGCTTGTGA	16242
GTAGGTGGGA	CTACAGGTG	ATGGGCGAT	GTCTGCTTAA	TTTTTGTTG	TTTTTACTAG	16302
AGACAGGGTT	TTGGTATGTT	GTCAGGCTG	GTCTTGGACT	GCTGATCTCA	GATGATCCTC	16362
CTGGCTCGG	CTCCCAAAGT	GCTGGGATA	CAGGATGAA	CCACGACACA	TGCGCTAAA	16422
ATTGATTCTT	ATGATTAATC	TCCTGTGAA	AAITGCTT	CATTTGAAAG	TTTGCCTTCA	16482
TTTGAAACCT	TCATTAA	GGCTGAGGAA	CAAGATGAGA	CCGCACTCT	ACAAAAAAAT	16542
GCAAAATATC	CTGTGAGAC	CTCTAATCT	CTGTTGAGGC	TGAAAGTAA	GGATCACTTG	16602
AGCTTAGGAA	TTTGAGCTG	CACTGAGGTA	TGATCCTGAA	GCTACACTCC	AGCTGCTG	16662
ACAGTAGAAC	CTGACACACA	CACACAAAAA	AAACCTTCA	TAAAAAAATTA	TTAGTTGACT	16722
TTTCTTAGGT	GACTTCCGT	TTAACGAAATA	AAATTAAAG	AAATATCTCT	AATTTTAGAA	16782
AATTATTTT	TAGTTACATA	TTGAATT	AAACGCTAG	TTTAAGTT	TATGTC	16842
TTACCTGAGA	ACACACTAAG	TCTGATAAGC	TTCAATT	GGCGCTTGT	GATGATTATA	16902
TAATATTCTG	ATGAAAGCCA	AGACAGACCC	TTAAACCATA	AAAATAGGAG	TTCGAGAAAG	16962
AGGAGTAGCA	AAAGTAAAG	CTAGAATGAG	ATTGAATTCT	GAGTCGAAT	ACAAAATTTT	17022

ACATATTCTG TTTCTCTT TTTCCCCCTC TTAG CT GAA GAT GAT G GTAAAGT 17075
 Ala Glu Asp Asp Glu

-10

AGAAATGAAT TTATTTTCT TTGCAAACTA AGTATCTGCT TGAGACATAT CTATCTCAGC 17135
 ATTGTCAGCT GAGGAAAAAA AAAATGGTT CTGATGCTAC CAATCTGCT TCAAAGAAAT 17195
 GTGGACTCAG TAGCACAGCT TTGGAAATGAA GATGATCATA AGAGATCAA AGAAGAACCT 17255
 CTAGCAAAAG ATGCTTCTCT ATGCGTTAAA AAATCTCTCA GCTTTAGAA TCTACAAAT 17315
 AGAGCTTGC TGTTTCATTG GTCTTAAGAT TAGCATGAAG CGATGGATTG TGTGTTAGGG 17375
 GGAGCGTTGC ATAGGAAAAA GGGATTGAGG CATTAGAATT GTCCAAAATC AGTAACACCT 17435
 CCTCTCAGAA ATGCTTGGG AAGAAGCCTG GAAGGTTCTG GCTGGGTGGT GGGGTGGGC 17495
 AGAAAATTCT GGAAGTAGAG GAGATAGGAA TGGGTGGGC AAGAAGACCA CATTAGAGG 17555
 CCAAAAGCTG AAAGAAACCA TGCCATTAT GATGAATTCA GGGTAATTCA GAATGGAAGT 17615
 AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGTAT AAAAGGGGTG TAGAGCAAGA 17675
 CGTTCTCTCA CCCCAAGATG TGAAATTCTG ATCTTATCTT GAGATAATA GGGTTAATT 17735
 AGCACAAATAT GTATTAGCTA GGTTAAAGAT TAGTTTGTG TAACAAAGAC ATCCAAAGAT 17795
 AGAGTAGCTG AATAAGATAG AGAATTCTG TCTCAAAGAA AGTCTAGTA GGAGCTCAG 17855
 AAGTAGTATG GCTGAAAGCA ACCTGATGAT ATTGGGAGCC CCAACCTCTC TCAGTCTTGT 17915
 ACCCATCATE CCCTAGTTGT TGATCTCTC ACATAGTTG AAAATCATCA TACTTCCTGG 17975
 GTTCATATCAG TTGTTATCAA GAAAGGCTCA AGAGAAGTCA GGGTCTCC TTTCAAAGAC 18035
 TCTAATTGGA AGTAAACAC ATCAATCTC CTCATATTCC ATTGACTAGA ATTTAATCA 18095
 ATGGCCACAC CAAGTCAAG GAAATCTGAA AAATATAATC TTTATTCCAG GTAGCCATAT 18155
 GACTCTTAA AATTCAAGAAA TAATATATTT TTAAATATC ATTCTGGCTT TGGTATAAAAG 18215
 AATTGATGTT GTGGGGTGGAG GAGGCGAAA TTAAGGGTTG AGAGCCTATT ATTTAGTTA 18275
 TTACAAGAAA TGATGGTGTG ATGAATTAAG GTAGACATAG GGGAGTGCTG ATGAGGAGCT 18335
 GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAGGAC ATGATTTAGG GTTGGATTG 18395
 GAAAGGAGAA GAAAGTAGAA AAGATGATGC CTACATTTT CACTTAGGCA ATTTGTACCA 19455
 TTCACTGAAA TAGGGAAACAG AGGAGGAGA GCAGGTTTG GTGTATACAA AGAGGAGGAT 18515
 GAGTGAACCA TTTCGTTTG GATCTGAGAT GTCTGTGGAA CGTCTAGTG GAGATGTCCA 18575
 CAAACTCTTC TACATGTGCT TCTGAGTTCA GGACACAGAT TTGGGCTGGA GATAGAGATA 18635
 TTGTTAGGCTT ATACATAGAA ATGGCATTG AATCTATAGA GATAAAAAGA CACATCAGAG 18695
 GAAATGTGTA AAGTGAAGAGA GCAAAAGGCA AGTACTCTGC TGGGGGAAAT ACCTACATT 18755
 AAGGGATGCA GTAGAAAGAA GCTAAATAAAC AACAGAGAG AAGTACCCA AAAGGGGAGA 18815
 AGAAAACCA AGAGAATTCC ACCGACTCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG 18875
 TGGTGTGTTG AATTCTGCAG CTTGAGAAAT CAAGGGCCAG CAACAAGGAG TTTGGTGTGATC GTGAAATGGA GGAGAGGCGA 18995
 GATTGCAATG AGTAAACAG TGAATGGAA GTGAAGAAAT GATACAGATA ATTCTGCTA 19055
 AAGCTTGGC TGTAAAGGAG AGGAGAGAAA CAAGACTAGC TCGAAAGTGA GATTGGATTG 19115
 ATGGAGGAGT TTTAAATCTC AAAATAAGA GCTTGTCT TTTTGATTA TGAAGATAT 19175
 GTGTTAATTG TAACTAATTG AGGCAATGAA AAAAGATAAT AATATGAAAG ATAAAATAT 19235
 AAAACACAG CAGAAATAAT GATAGCTTAC ATTCTGATAC AATATTTCTA CACTCCTTTC 19295
 TATGTATATA TACAGACACA GAAATCTTA TATTTTTATT AAAAGGGATT GTACTATACC 19355
 TAACTCTT TTTCTAGTTA GTGATATATA TGGACATCTC TCGATGCAA CGAGTAATTG 19415
 CAGTTATATT AAGTCATCA TATTTCACAA TAAGGGCATA TCTTGGCCCT TTTTATTAA 19475
 TCAATTCTTA ATTGGGTGAAT GTTTGGTCTC AGTTGTTGT TGTTATTAAAC AATGTTCCCA 19535
 TAAGCTTCC TGTACACCAA TGTTCACACCA TTTGTCGTGAT TTTTCTCTCA GGATAAAAC 19595
 CAGGAGGTTAG AATTGCTGGG TTGATAGAAG AGAAAGGATG ATTGOCAAAT TAAAGCTTCA 19655
 GTAGAGGAGTA CATGCCAGC ACAAATGGGAA TCAAGCCCTAG ATACCAGAAA TGGCACTTTC 19715
 TCATTTCCCT TTGGGACAAA AGGGAGAGAG GCAATAACTC TGCTGCCAGA GTAAATTG 19775
 TAAGTGGAGT AGCAGGAAAT CATTTCCTGA AATATGAAAC AGAGATGATG TTGTTAGAGGT 19835
 CTGAAAGAGA GCAAAGAAAA TTGAAATTG CGGCTATCAG CTATGAAAGA GAGTGTGAA 19895
 CTGGAAAAAA AAAGAAAGTAT TGCACATTG TATGCTTGTG ATGGCAGCGA TTGAAAGGCT 19955
 TGTGCCATCTG TTGACACAGCA GCACCTCAGCA GCGAACATTG GAGTTTGTA GCAAGAAAGAC 20015
 AAATAAGTTA GGGATTAAAT ATCCTCTCCTA AATGCTGAGC AAAATGAACT CTGAGATCCA 20075
 GCTGACAGC GAAGGAAAGGG AAGACGGGAA GAGGGTAGAT AGGAAATACA AGAGTCAGGA 20135
 GACTGGAAGA TGTGTTGATA TTGAAAGACA CATAGAGTT GAGTAAAGT GTAAAGAAAAC 20195
 TAGAAGGGTA AGAGACCGGT CAGAAAGTAG GCTATTTGAA GTTAACACTT CAGAGGCGAGA 20255
 GTAGTTCTGA ATGGTAAACAA GAAATTGAGT GTGCTTCTGAGTAAAGTTA AAAAACAAATA 20315
 GCGAACTTTA TTGTTAGCTAC TTCTGAAACCA GAAGATTGTC ATTAATAGTT TTGAAAAACT 20375
 AAAATATATA GCATACTTAT TTGTCATTA ACAAAAGAAAC TATGTTTTT TAAATGAGAT 20435
 TTAATGTTA TTGTTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA 20486

Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu

-5

1

5

TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT 20534
 Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile

10

15

20

GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT 20582
 Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys

AGA G GT ATTTTTTTA ATTCGAAAC ATAGAAATGA CTAGCTA CTT CTTCCATTG 20638

Arg Asp

40

TGTTTACTG CTTACATTGT TCCGTGCTAG TCCCAATCCT CAGATGAAAA GTCACAGGAG 20693
 TGACAATAAT TTCACTTAA GAAGAACTTTA TAAGGCATCC ACCTTTTTA GTTGGGGTAA 20758
 AAAATTGGAT ACAATAAGAC ATTCGCTAGGG GTCATGCCTC TCTGAGCTG CCTTTGATC 20818
 ACCAATCCCT TTATTGTGAT TGCACTTAACT GTTTAAAACC TCTATAGTTG GATGCTTAAT 20878
 CCCTGCTTGT TACAGCTGAA AATGCTGATA GTTTACCAAGG TGTGGTGGCA TCTATCTGTA 20938
 ATCCTAGCTA CTTGGGAGGC TCAAGCAGGA GGATTGCTTG AGGCCAGGAC TTTGAGGCTG 20998
 TAGTACACTG TGATCGTATC TGTGAATAAG CACTGFACTC CAGCTGGGT GATATACAGA 21058
 CCTTGTCTCT AAAATTAAAAA AAAAAGGAAAGGAAATTGATCAAG 21118
 TCTACTGTGCTT CTTCAAAAC ATGAATTCCA AATATCAAAG TTAGGCTGAG TTGAAAGCTAGT 21178
 GAATGTGCAT TCTTTAAAAA TACTGAATAAC TTACCTTAAAC ATATATTTA AATATTTAT 21238
 TTAGCATTAA AAAGTTAAAA AGCAATCTTT AGAATTCATA TCTTTAAAAT ACTCAAAAAA 21298
 GTTGCAGCGT GTGTGTGTA ATACACATTA AACTGTGGGG TTGTTGTTT GTTGTGAGATG 21358
 CAGTTTCACT CTGTCACCCCA GCGTGAAGTG CASTGCAGTG CAGTGGTGTG ATCTCGCTC 21418
 ACTACAAACCT CCACCTCCCA CGTCAAGGG ATTCTCATGC CTCAGTCTCC CGAGTAAAGTG 21478
 GGATTACAGG CATGCACCCAC TTACACCCGG CTATTTTTG TATTTTTAGT AGAGCTGGGG 21538
 TTTCACCATG TTGGCCAGGC TGGTCTCAAA CCCCTAACCT CAAGTGTCT GCCTGCCTCA 21598
 GCCTCCCAA CAAACAAACA ACCCCACAGT TTAATATGTG TTACAACACA CATGCTGCAA 21658
 CTTTTATGAG TATTTTAATG ATATAGATTA TAAAAGGTTG TTTTTAACTT TTAAATGCTG 21718
 GGATTACAGG CATGAGCCAC TGTGCCAGGC CTGAACGTG TTTTTAAAAT TGCTGACCA 21778
 GCTGTACATA GTCTCGCA GACTGGCCAA GTCTCAAAGT GGGAACAGGT GTATTAAGGA 21838
 CTATCCTTTG GTTAAATTTG CGCAAATGTT CCTGTGCAAG AATTCTCTA ACTAGAGTTG 21898
 TCATTTATTA TATTTATTTG AG AT AAT GCA CCC CGG ACC ATA TTT ATT ATA 21949

Asp Asn Ala Pro Arg Thr Ile Ile Ile

45

AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT 21997
 Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile Ser
 50 55 60 65

GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT 22045
 Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile
 70 75 80

TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTCAAT CATGTTAATA TATCTAATAT 22103
 Ser Phe Lys
 AATTAGAAAT ATAACATTAT TTCTAATGTT AATATAAGTA ATGTAATTAG AAAACTCAA 22163
 TATCCTCAGA CCAACCTTT GTCTAGAACCA GAAATTACAA GAGCAGAGA ACCATTAAAG 22223
 TGAATACTTA CTAAAAATTAA TCAAAACTCTT TACCTATTGT GATAATGATG GTTGTCTGA 22283
 GCCTGTCACA GGGGAGAGGG AGATACAAACA CTTGTTTAT GACCTGCTATC TCTGAAACAA 22323
 TCAGTCTTTA TACAATAAT AATGTAGAAT ACATATGTG GTTATACATT TAAGAATAAC 22403
 ATGTGACTTT CCAGAATGAG TTCTGCTATG AAGAATGAAG CTAATTATCC TTCTATATTT 22463
 CTACACCTTT GTAAATTATG ATAATATTTT AATCCCTAGT TGTTTGTG CTGATCCTTA 22523
 GCGTAAGTCT TAGACACAAAG CTTCAGCTTC CAGTGTGATGT ATGTTATTTT TATGTGTTAT 22583
 CTAATTGAAT AAAAGTTATG AGATCAGCTG TAAAGAATAT GCTATAPTTA TCTTCAAGGCC 22643
 AGGTATAAG TATTTCTGGC CTCTACTTT TCTCTATTAT TCTCCATTAT TATTCCTAT 22703
 TATTTTCTC TATTTCTCC ATTATGTTA GATAAACCA AATTAACCTAT AGCTACABAC 22763
 TGAGCCAGTA AGAGTAGCCA GGGATGCTTA CAAATTGGCA ATGCTTCAGA GGAGAATTGC 22823
 ATGTCATGAA GACTCTTTT GAGTGGAGAT TTGCGAATAA ATATCCGCTT TCATGCCAAC 22883
 CCAGTCCCCA CTGAAAGACA GTTAGGATAT GACCTTAGTG AAGGTACCAA GGGGAAATT 22943
 GGTAGGGAGA AAAAAGGAG TCTAAAATAT AATCCTAAAGTA AGAACATGTC ATATGAAACA 23003
 GATACAGCCC CCAGACAAAT CCCTCAGTTA TCTCCCTCCA ACCAGACTGC CACCCCTTCA 23063
 GSTGAAATT TGGAGTCCC ATTCTAGGCC TGACAGGCAG CTTAGTTATC AAAATAAT 23123
 AAGAGGCTG GGTGAGGG GTAGGGTGG AAGGGTTAAG CATGCTGTTA CTGAAACATA 23183
 TAATTAAGAAG GGAAGGAGAT GCGCAAGTTC AAGCTATGTG GGATAGAGGA AAAATTCACT 23243
 GAGAGGAGC ATTCAAGAAC TGGGATAAGT CGAAAGCTAC AGGTGGATTC TTGTTGAAAGG 23303
 AGACTGGTGA AAATGTTAAG AAGATGGAAA TAATGCTTGC CACTTATGAG GAAGTGGGCA 23363
 AATCCATATT TGGGGGAGGC TGAAGTTAT TCAATTGTA TGCGCCTTTT AAAATAAAAAG 23423
 AATGTGGCTG GCGCTGCTGG CTCACACCTG TAATGCCAGC ACTTTGGGAG GCGGAGGGGG 23483
 GCGGATCAGC TGAACTGAGG AGTTCAAGAC CAGCTGACG AACATGAGA AACCCCATCT 23543
 CTACTAAAAA TACAAAATTA GTGCGGCGTG GTGGCATATG CCTGTAATCC CAGCTACTCG 23603
 GGAGGCTGAG GCAGGAGAAT CTTTGAACCC CGGGAGGCGAG AGGTGEGAT GAGGCTAGAT 23663
 CGTGCCTATG CACTCCAGGC TGGGAAACAA GAGCAAAACT CGGTCTCAA AAAAAGAAAAA 23723
 AAAAAGTGAATTAACCAAGG GCGATTAGCT TAATATTAA ATACTGTTTT TAAAGTGGGC 23783
 GGGGGGTGGC TGGAAAGAGAT CTGTGTAAT GAGGGAATCT GACATTAAAG CTTCATCAGC 23843
 ATCATAGCAA ATCTGCTTCT GGAAGGAAC CAATAAAATAT TAGTTGAGG GGGGAGAGA 23903
 GTGAGGGGTG GACTAGGATC AGTTTGTAGGC CTTGTCTTTA ATCCCTTTG CTGCCACTAA 23963

TAAGGAT-CTT AGCAGTGGTT ATAAAAGTGG CCTAGGTTCT AGATAATAAG ATACAACAGG 24023
 CCAGGACAG TGGCTCATGC CTATAATCCC AGCACTTGG GAGGGCAAGG CGAGTGTCTC 24083
 ACTTGAGATC AGGAGTCAAA GACCAAGCCTG GCGAGCATGG CGATACTCTG TCTCTACTAA 24143
 AAAAATACAA AAAATTAGCC AGGCATGGTG GCATGCACCT GTAAATCCCAG CTACTCGTGA 24203
 GCCTGAGGCA GAAGAATCGC TTGAAACCAAG GAGGTGTAGG CTGCAGTGTAG CTGAGATGCG 24263
 ACCACTGCAC TCCAGCTGG GOGACAGAAT GAGACTTTGT CTCAAAAAAA GAAAAAGATA 24323
 CAAACAGGCTA CCCTTATGTG CTCACCTTTC ACTGTTGATT ACTAGCTATA AAGTCCTATA 24383
 AAGTTCTTGT GTCAGAACCC TTGACAAACAC TAAGAGGGAT TTGCTTTGAG AGGTTACTGT 24443
 CAGAGTCTGT TTCATATATA TACATATACA TGTATATATG TATCTATATC CAGGCTTGGC 24503
 CAGGGTTCCC TCAGACTTTC CAGTGCACCT GGGAGATGTT AGGTCAATAT CAACTTTCCC 24563
 TGGATTCAAGA TTCAACCCCT TCTGATGTAA AAAAAGAAAA AAAAAGAAAA GAAATCCCTT 24623
 TCCCCTTGGGA GCACCTCAAGT TTCACCAAGGT GGGGCTTCC AAGTTGGGG TTCTCCAAGG 24683
 TCATTGGGAT TGCTTTCACA TCCATTGCT ATGTACCTT CCTATGATGG CTGGGAGTGG 24743
 TCAACATCAA AACTAGGAAA GCTACTGCC AAGGATGTGC TTAACTCTAT TCTGAAATGT 24803
 GCAATAAATG TGATTAAGA GATTGCTGT TCTACCTATC CACACTCTCG CTTTCAACTG 24863
 TAACCTTCTT TTTTCTTTT TCTTTTTTT TGAAAGGGAG TCTCGCTCTG 24923
 TCGCCCAAGGCG TAGAGTGCAG TGGCACGATC TCAGCTCACT GCAAGCTCTG CCTCCCGGGT 24983
 TCAAGGCTATT CTCCTGCCTC ACGCTTCCAA GCAGCTGGGA CTACAGGCGC CTGECACCAT 25043
 GCCCAGCTAA TTTTTGTAT TTTTAGTAGA GAGGGGTTT CACCGTGTAA GCCAGGATGG 25103
 TCTCGATCTC CTGAACCTGT GATCCCGCCG CTCAGGCTC CCAAAGTGCT GGGATTACAG 25163
 GCGTGAGCCA TCGCACCCGG CTCAACTGTA ACTTTCTATA CTGGTTCATC TTCCCTGTG 25223
 ATGTTACTAG AGCTTTGAA GTTTGGCTA TGGATTATTT CTCACTTATA CATTAGATTT 25283
 CAGATTAGTT CCAAATTGAT GCCCACAGCT TAGGGTCTCT TCTAAATTG TATATTGTAG 25343
 ACAGCTGCAAG AAGTGGGTGC CAATAGGGGA ACTAGTTTAT ACTTTCTATCA ACTTAGGACC 25403
 CACACTTGTGTT GATAAAGAAC AAAGGTCAG AGTTATGACT ACTGATTCCA CAACTGATTG 25463
 AGAAGTTGGA GATAACCCCG TGACCTCTG CATCCAGAGT CTTTCAGGCA TCTTGAAGG 25523
 ATGAAGAAAT GCTATTAA TTTTGGAGGT TTCTCTATCA GTGCTTGTAGA TCATGGAAAT 25583
 CTGCTGCTGCC ATGAGGCCAA ATTAAGTCC AAAACATCTA CTGCTTCCAG GATAACATG 25643
 GAAACACCTT AGGTGGTGC CACATGTTCT GATCGATCT GCAAGATAGA CATGCTGCAC 25703
 TAACAGGAAA AGTGCAGGCA GIACTACAG TGGATAACC TCGAGGATTA TAGTTCAAG 25763
 TAATCTAACCT ATTTCACCA AGGCCCTATT CTGTGACTGA AACATACAAG AATCTGCTT 25823
 TGGCCTTCTA AGGCAGGGCC CAGCCAGAGA GACCATATTC AGGACAGAAA TTCAAGACTA 25883
 CTATGAACT GGAGTGCTTG GCAGGGAAAGA CAGAGTCAG GACTGCCAAC TGAGCCAAATA 25943
 CAGCAGGCTT ACACAGGAAC CCAGGGCTA GCGCTACAA AATTATGGG TCTATTCACT 26003
 GTAAGTTTA ATTTCAGGCT CCACTGAAAG AGTAAGCTAA GATTGCTGGC ACTTTCTGTC 26063
 TCTCTCACAG TTGGCTCAGA ATGAGAAACT GGTAGGCCA GCGATGCTGG CTTACACCTG 26123
 GAATCCCAGC ACTTTGGGAG GCGGAAGTGG GAGGGTCATT TGAGGCCAGG AGTCAGGAC 26183
 CAGCTTAGGC AACAAAGTGA GATACCCCT GATCCCTTCT CTAAAAAAAT AAATTCTAA 26243
 AATTAGCCAA ATGTGGTGGT GTATACTTAC AGTCCTCAGT ACTCAGGAGG CTGAGGCCAGG 26303
 GGGATTGCTT GAGCCCAGGA ATTCAAGGCT GCACTGAGT ATGATTCAC CACTGCTT 26363
 CTGGCTGGGC AACAGAGCGA GACCCCTGTCT CAAAGCAAA AAAAAAGAA ACTAGAACTA 26423
 GCCTAAGTTT GTGGGAGGAG GTCATCATCG TCTTAGCGTG TAAATGGTTA TTATAGAGGA 26483
 CAGAAATTGA CATTAGCCCC AAAAGCTTGT GGTCTTTGCT GAAACTCTAC TTAATCTGA 26543
 GCAAATGTGG ACACCACTCA ATGGGAGAGG AGAGAACTAA GCTGTTGTAT GTATAAGGG 26603
 AAACTAAGGG CCTGAAACTG AATATGATC CCATGACAGG GAGAATAGGA GATTGAGGAGT 26663
 TAAGAAGGGAG AGGAGGTCAG TACTGCTGTT CAGAGATTT TTTTATGTAA CTCTTGAGAA 26723
 GCAAAACTAC TTTTGTCTG TTGCTAATA TACTCAAAA CAAACTTCAT ATATTCAATT 26783
 TGTTCATGTC CTGAAATAAT TAGGATATCT TTTTTCTCT ATAG GAA ATG AAT CCT 26839
 Glu Met Asn Pro
 85

CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA 26887
 Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg
 90 95 100
 AGT GTG CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC 26935
 Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr
 105 110 115 120
 GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC 26983
 Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu
 125 130 135
 ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT 27031
 Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr
 140 145 150
 GTT CAA AAC GAA GAC T AGCTATTAAA ATTTCATGCC GGGCGCAGTG GCTCACGCC 27087
 Val Gln Asn Glu Asp
 155
 GTAATCCCAG CCCTTGGGA GGCTGAGGGC GGCAGATCAC CAGAGGTCAG GTGTTCAAGA 27147
 CCAGCCTGAC CAACATGGTG AACACCTCATC TCTACTAAAA ATACAAAAAA TTAGCTGAGT 27207

GTAGTGACCC	ATGCCCTCAA	TCCCAGCTAC	TCAAGAGGCT	GAGGCAGGAG	AATCACTTGC	27267
ACTCCGGAGG	TGGAGGTTGT	GGTGAGCCGA	GATTGCACCA	TTGCGCTCTA	GCCTGGCAA	27327
CAACAGAAA	ACTCCATCTC	AAAAAATAAA	ATAAAATAAT	AAACAAATAA	AAAATTCTATA	27387
ATGTGAACTG	TCTGAATTTC	TATGTTAGA	AAGATTATGA	GATTATTAGT	CTATAATTGT	27447
AATGGTAAA	TAAAATAAT	ACCAGTCTTG	AAAAACATCA	TTAAGAAATG	AATGAACCTTT	27507
CACAAAAGCA	AACAAACAGA	CTTCCCTTA	TTTAAGTGA	TAAAATAAAA	TAAAATAAAA	27567
TAATGTTAA	AAAATTCTATA	GTGAAAAC	ATTCTACATT	TTAATTGGC	ATATTAATTA	27627
TACTTAATAT	AATTATTTT	AAATCTTTG	GGTTATTAGT	CCTAATGACA	AAAGATATTG	27687
ATATTGAAAC	TTTCTAATT	TTAAGAATAT	CGTTAAACCA	TCAATATTTC	TATAAGGAGG	27747
CCACTTCACT	TGACAAATT	CTGAATTTC	TCCAAAGTCA	GTATATTTC	AAAATTCACT	27807
TTGATCCTGA	ATCCAGCAAT	ATATAAAAGG	GATTATATAC	TCTGGCCAAC	TGACATTCTAT	27867
CCTAGGAATG	CAAAGATGGT	TTAATATCCT	AAAATCAATT	AACATAACAT	ACTATATTAA	27927
TAAAGTATCA	AAACAGTATT	CTCATCTTT	TTTCTTTTT	CACAATTCC	TGGTTACACT	27987
ATCATCTCAA	TAGATGCAGA	AAAAGCATT	GACAAATCC	AATTCTATAAT	AAAATTCTC	28047
AAACTGAAA	GAGAACATCA	TAAGGCATC	TATGAAAAAC	CTACAGCTAA	TATCCTACTT	28107
AACGATGAAA	AACTGAATTA	TTTACCTCA	AGATCAAGAA	TAATGCAAGC	ATGTCAGCTC	28167
TTGCAACTTC	TATTCAACAT	TGTACTGGAG	GTTCTAGCCA	GAGCAACCAT	ACAATAAATA	28227
AAAATAAAAG	GCAECCAGAT	TAGAAAGGAA	GTCTTTATT	GCAGACAACA	TGGTTCTTTA	28287
TGCAGAAC	CGTCAGGAAT	ACACACACAT	GTTAGAACTA	ATAAGTTCA	CAAGGTTGCA	28347
GGTGCAATA	TCAATATGCA	AAATACATT	GAAGGCTGGG	CTCAGTGGAG	ATGGCATGTA	28407
CCTTCGTCC	CAGCTACTTG	GGAGGCTGAG	GTAGGAGGAT	CACTTGAGGT	GAGGAGTTG	28467
AGGCTATAGT	GCAATGTGAT	CTTGCCTGTG	AATAGCCACT	GCACCTGAGC	CTAGGCAACA	28527
AAGTGAGACC	CCGTCTCCAA	AAAAAAAAT	GGTATATTGG	TATTTCTGTA	TATGAACAAAT	28587
GAATGATCTG	AAAACAAGAA	ATTCCATT	ACGATGGTAT	AAAAAAATA	AAATACAAAT	28647
AAATTAGCA	AAATAATTAT	AACTGTGAT	CATCGAAAAT	TTCAAAGCAC	TCTGAGGGAA	28707
ATTAAGATG	ATCTAAATAA	TTGGAGAGAC	ACTCTATGAT	CACTGATGG	AAAATTCTATT	28767
CAATATTGTT	AAGATAACAA	TTGTCCCCAA	ATTGATGCAT	GCATTCAATT	TAGTCTTCAT	28827
CAAAATTCCA	GCAGGGTTT	TGAGAAATT	GACAAGCTGT	ACCCAAATG	TATATGGAAA	28887
TGAAAEGRCC	CAGAAGAGCA	ATAATTTC	AAAAACAAA	GTTGGAAAAC	TTTACTTC	28947
TAATTTAAA	ACTTACTATA	AACTAAAGT	TATCAAGACC	ATTTAGT		28994

(15) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: N-terminal fragment
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
 1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCATCCTAAT ACGACTCACT ATAGGGC

27

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTATAGGGCA CGCGTGGT

18

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAGTTTTC ACCTTCCAAC TGTAGAGTCC

30

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGGATCAAGT CGTGATCAGA AGCAGCACAC

30

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCTGGCTGCC AACTCTGGCT GCTAAAGCGG

30

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTATTGTCAA TAAATTCAT TGCCACAAAG TTG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAGATGGCTG CTGAACCAAGT AGAAGACAAT TGC

33

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCCTTGGTCA ATGAAGAGAA CTTGGTC

27

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCTGGAAATCA GATTACTTTG GCAAGCTTGA ATC

33

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCAGCCTAG AGGTATGGCT GTAATATCT C

31

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATGAAAT TTTAATAGCT AGTCTTCGTT TTG

33

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGACATCAT ATTCTTCAG AGAAGTGTC

30

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAATTGAA TCTTCATCAT ACGAAGGATA C

31

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCCGAAGCTT AAGATGGCTG CTGAACCA GT A

31

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAAATAATT TTGTTCTCAC AGGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ATGTAGCGGC CGCGGCATGA AATTTAATA GCTAGTC

37

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGGAAATCA GATTACTTTG GCAAGCTTGA ATC

33